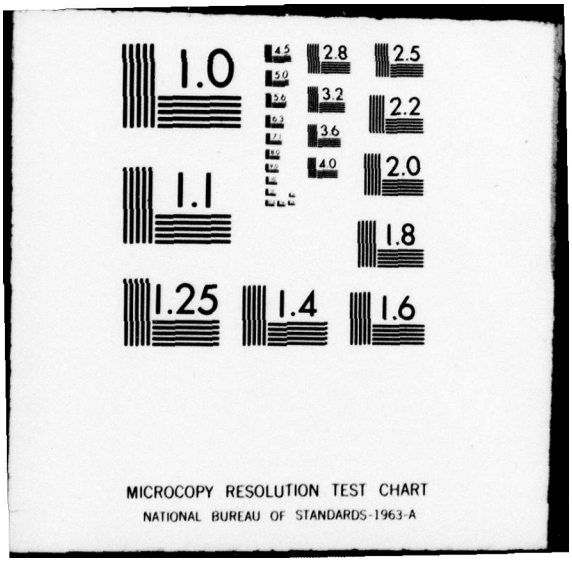
AD-A065 489 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO F/G 5/10
THE CLOZE TEST AS A PROCEDURE FOR ESTABLISHING OBJECTIVE GERMAN--ETC(U)
1978 B M STRONG
UNCLASSIFIED AFIT-CI-79-100D NL

1 OF 3
AD
A065 489


1.0
1.1
1.25
1.4
1.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.8
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

79-100D
vol 1
LEVEL II

THE CLOZE TEST AS A PROCEDURE FOR ESTABLISHING OBJECTIVE GERMAN PROSE READABILITY STANDARDS.

DISSERTATION

Doctoral thesis,

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

236 p.

By

Brent M. Strong, B.A., M.S.

* * * * *

AFIT-CI-79-100D

The Ohio State University

1978

Reading Committee:

Professor Edward D. Allen
Professor Gilbert A. Jarvis
Professor Victor M. Rentel

Approved By

Edward D. Allen
Adviser
College of Education

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER CI 79-100D	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* The Cloze Test as a Procedure for Establishing Objective German Prose Readability Standards		5. TYPE OF REPORT & PERIOD COVERED Dissertation 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Brent M. Strong		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS AFIT student at Ohio State University		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS AFIT/CI WPAFB OH 45433		12. REPORT DATE 1978 13. NUMBER OF PAGES 225
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for Public Release, Distribution Unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

JOSEPH P. HIPPS, Major, USAF
Director of Information, AFIT

APPROVED FOR PUBLIC RELEASE AFR 190-17.
FEB 8 1979

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

To My Parents

On Their

50th Wedding Anniversary

ACKNOWLEDGMENTS

I would like to express my appreciation to my reading committee, Professors Edward D. Allen, Gilbert A. Jarvis, and Victor M. Rentel. From each I received invaluable advice, assistance, and encouragement in conceptualizing and conducting this research.

Additionally, I would like to thank Professor Werner Haas of the German Department for his assistance in selecting passages, proofreading tests, and paving the way for obtaining permission to collect the data.

To my Air Force colleague, Captain Michael D. Bush of the United States Air Force Academy, I am especially indebted. His suggestions and help in both using the computer and analyzing the findings contributed significantly.

Finally, I express my gratitude to my wife, Kelly. Without her love, understanding, and willingness to shoulder additional family responsibilities in raising our six children, this undertaking could never have been accomplished.

VITA

September 6, 1939	Born - Idaho Falls, Idaho
1963	B.A., Brigham Young University, Provo, Utah
1964-1972	Intelligence Officer, United States Air Force
1974	M.S., Georgetown University, Washington, D.C.
1973-1974	German Instructor, Department of Foreign Languages, United States Air Force Academy, Colorado
1975	Assistant Professor of German, Department of Foreign Languages, United States Air Force Academy, Colorado
1975-1977	Air Force-Sponsored Doctoral Candidate, Department of Humanities Education, The Ohio State University, Columbus, Ohio
1978	Chairman of Basic Courses, Department of Foreign Languages, United States Air Force Academy, Colorado

FIELDS OF STUDY

Major Field: Foreign Language Education

Studies in Methodology, Research, Supervision, and TESOL. Professor Edward D. Allen

Studies in Educational Psychology, Research, Statistics, and Testing. Professor Gilbert A. Jarvis

Minor Field: Curriculum and Foundations

Studies in Instruction and Interaction Analysis.
Professors J. B. Hough and J. K. Duncan

Studies in Curriculum. Professor P. R. Klohr

TABLE OF CONTENTS

TABLE OF CONTENTS-CONTINUED

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES-CONTINUED

CHAPTER I

INTRODUCTION

Rationale

The Importance of Reading

Of all the skills one acquires as a result of formal education, the ability to read must certainly be ranked as one of the most important. Literacy, which is broadly defined by Bormuth (1973-1974) as nothing more than the ability to respond appropriately to written language (the ability to read) always has been given prominent emphasis by various social movements seeking to improve the status of mankind. Reading has received such emphasis because of an acute awareness that little in life of any great value is ever accomplished without in some way involving reading.

Because of its great importance, the active development of reading is begun in children shortly after they enter elementary school. Initially, students read primarily for practice. Very soon, however, as more and more knowledge is transmitted through the medium of written language, academic survival becomes primarily dependent upon the ability to read.

The Importance of Foreign Language Reading

As in all other academic disciplines, reading also plays a central role in the learning of a foreign language. Although oral skills frequently are emphasized initially in many foreign language programs today, primarily as a result of the influence of the audio-lingual method, which gained so much popularity in the 1960s, Phillips (1974) points out that nevertheless, relatively quickly students begin to devote considerable time to reading. For example, even while speaking skills are being given the dominant emphasis, speaking practice in class often centers around assigned reading passages. Beyond the beginning level, of course, various reading activities receive a major portion of the student's attention. So important is reading to the overall foreign language classroom environment, that Phillips (1974) indicates that its elimination "would make it extremely difficult to teach any language skill in the traditional school setting" (p. 1). Ultimately, then, being a successful second-language reader most likely determines student achievement, in the total foreign language program, more than any other skill.

Because of the centrality of reading, it would therefore seem most apparent that any pedagogical practice or procedure designed to assist in either refining, improving, or evaluating student reading skills would be welcomed by both teachers and students alike.

Problem Statement

Reading and Comprehension

For years educators have struggled with the problem of defining what skills are involved in reading. Over 30 years ago, Davis (1944) pointed out that there was no widespread agreement as to exactly what reading ability was. Today, there is still no widespread agreement. Authorities generally agree only that reading consists of a collection of different skills. Some, however, have postulated large numbers of skills, as for example Burkhart (1945), who identified 214 separate skills; while others have preferred to define reading in terms of a relatively small number of skills. Anderson (1949), for example, using statistical techniques, limited reading to just three factors.

Although educators and reading researchers over the years have been unable to agree upon exactly which skills are involved in reading, there nevertheless has been general agreement that the student must be able to understand what he or she reads. Bormuth (1969) indicates that the ability to comprehend is one of the most basic factors determining the effectiveness of instruction. He says that if the student is unable to understand the language of the school's curriculum, "he will almost certainly fail to learn much of the content of his instruction and both the student and the instruction will fail to attain their objective" (p. 3).

In discussing foreign language reading, Allen and Valette (1972) suggest that reading involves more than just assigning foreign language sounds to the written words. Instead, "it requires the comprehension of what is written" (p.189). Chastain (1971) points out that the ultimate goal in reading a foreign language is to be able to read comfortably:

> The word "comfortably" implies that the students should reach a level at which they do not feel a conscious strain while reading, i.e., a level at which they can concentrate on the message without being consciously aware of the code. Naturally, students who spend most of their time thumbing through a dictionary while preparing an assignment have not obtained a reading level consistent with the aforementioned objective (p. 177).

Although there is no widespread agreement as to exactly what skills are involved in reading, it is nevertheless evident that the student must be able to understand the reading material. In fact, in foreign language instruction, one of the key criteria in selecting reading passages ultimately must be whether or not the material can be comprehended. Because the reading skill is so central to successful foreign language learning, great care must be exercised by the teacher in choosing reading selections that are comprehensible.

Readability

One key factor that is often underemphasized in determining suitable reading selections is the difficulty level of the passage, i.e., its readability. Unfortunately, foreign language students are frequently introduced to

reading selections far too advanced for their reading capabilities. At the very least, such situations result in extreme frustration, which if unchecked, may lead to disinterest and possible disenrollment--a tragic occurrence under any circumstances.

In examining the phenomenon of reading difficulty, one quickly recognizes that it is a most complex problem influenced by a wide variety of factors. Anderson (1976) suggests that readability factors can be divided broadly into two categories: material factors and subject factors. For example, reading difficulty is undoubtedly influenced by such factors as vocabulary, the complexity of the sentence structures, the amount of redundancy present, the style employed by the writer, the format utilized, typographical considerations, etc., in other words, all those factors that are inherent in the material itself. Additionally significant, however, are those subject factors related directly to the reader himself such as intelligence, aptitude, attained reading level, motivation, experience, etc. That both of these categories of factors are very important in determining reading difficulty levels has been a well known fact for some time. Almost thirty years ago, Betts (1949) indicated:

> Reduced to its lowest terms, readability is a two-way proposition. The first consideration is the reader--his experience, his interests, his feelings, his motivation, his language facility, his needs, and his reading and study conditions. Any means of predicting readability is valid to the degree that the reader is taken into account.

> The second consideration is the interest level, the language, the mental constructs, and the mechanical features of the reading material (p. 438).

Most recently, the importance of the interaction between reader and materials has been stressed by Anderson (1976), who indicates that "any measure of readability should take these two factors (material and subject) into account" (p. 5).

Considering the complexity of accurately assessing readability, which clearly involves both material factors and subject factors, a most significant question is raised: How can teachers, especially foreign language teachers, know if reading selections they have chosen are of an appropriate difficulty level for their students?

English Readability Studies

The topic of assessing reading difficulty has been an area of interest and concern for many years. According to Lorge (1944), the first recorded attempts to apply readability procedures were made by the Talmudists, who, as early as 900 A.D., made word and idea counts to distinguish usual from unusual meanings. In addition to religious teachers, Gray (1937) indicates that educators began studying elements of reading difficulty centuries ago in connection with children's reading. Evidence of this interest appeared as early as 1840 when ease-of-understanding in the McGuffey Readers was considered in terms of vocabulary.

More modern studies of English readability date back to the early 1920s. Chall (1958) awards credit for the first quantitative study in readability to Pressey and Lively, who published a paper on vocabulary burden in 1923. Since that time, the problem of readability has received ever-increasing attention from educational researchers.

Most widely pursued has been the concept of the readability formula. Defined by Klare (1963) as "a method of measurement intended as a predictive device that will provide quantitative, objective estimates of the style difficulty of writings" (p. 3), the aim of the readability formula has been viewed as a fairly accurate yardstick for measuring the reading difficulty of a given text. Some of the more prominent early researchers who have developed popular English readability formulas include: Lorge (1939), Flesch (1943 and 1948), and Dale and Chall (1948).

Foreign Language Readability Studies

Regarding foreign language readability studies, Klare (1963) indicates that most research has been done in the United States. The earliest work detected by this researcher was a study conducted in 1927 by Johnson, who suggested that French could be rated according to difficulty by determining the number of "uncommon" words per 1,000 running words. In reviewing foreign language readability research up to 1950, Spaulding (1950) was able to uncover only a

handful of studies, all of which were conducted in either French or Spanish.

Concerning the development of actual foreign language readability formulas, only three were uncovered: 1) According to Hall (1971), the first readability formula devised specifically for foreign language materials was developed in 1939 by Tharp, who, by utilizing the factors of density and frequency, tried to determine the relative difficulty of French texts for English-speaking readers. 2) The most widely known foreign language readability formula was published in 1956 by Spaulding, who simplified a formula he had devised in 1951 for use in Spanish. Based on the Dale-Chall formula (1948), Spaulding's formula contained two variables: average sentence length and vocabulary density. 3) In 1963 Scherer modified and adopted Spaulding's formula for use on German texts intended for English-speaking readers.

Readability Formula Weaknesses

Although the idea of assessing readability has been a topic of concern for centuries, only during the last 50 years has the subject been seriously pursued. While only limited findings have been recorded involving foreign language reading difficulties, the bulk of the readability research has been accomplished in English and has centered around the development of the readability formula. Despite the fact that prominent researchers have developed popular formulas, however, difficulties have remained. In discussing

the problems of readability formulas, Bormuth (1967a) points out the following:

> Perhaps one of the chief reasons why instructional materials are not routinely evaluated to determine whether they have a suitable level of difficulty is that there has been no technique that is at once convenient, economical, and valid. Readability formulas are convenient, inexpensive, and require only unskilled clerical assistance to use, but the formulas presently available have validities that range from .5 to only about .7. Moreover, the equations take into account only a limited range of linguistic variables and the variables that are taken into account are, by today's standards, crude (pp. 2-3).

Why have readability formulas generally experienced validity difficulties? Two reasons are immediately apparent: First, as Bormuth suggests, only a limited range of linguistic variables has been considered, and these are rather crude by today's standards. The second reason is found in something Bormuth's statement implies. That using relatively crude linguistic variables to assess reading difficulty levels would decrease the valitity of a readability formula is readily understandable. More important, however, is the fact that only linguistic variables have been considered. As Anderson (1976) has indicated, the area of readability must be divided into both material factors and subject factors. Because traditional readability formulas have focused primarily on material factors and have almost totally neglected subject factors, it is little wonder that validity levels have ranged no higher than between .5 and .7. If in assessing reading difficulty, validity levels of readability

formulas are to be improved, additional factors must be taken into consideration.

Direct Testing, Another Possibility

If traditional readability formulas are incapable of accurately and validly assessing reading difficulty because their linguistic variables are crude, and because they fail to take into account subject factors, what other options are available? According to Bormuth (1975), "the time honored way to find out if a selection is suitable for students is to test them on it" (p. 62). To do this, the teacher identifies the appropriate reading selection, writes some questions over the material, has the students read the material and then try to answer the questions, and then scores the test. Finally, some criterion score, say 75 percent, is used to interpret student performance by supposedly representing a minimum level of desirable performance. For those who score at or above the criterion score, the material is considered readable and therefore suitable.

Direct-Testing Weaknesses

It is evident that direct reading tests take into account both material factors, because the passage of concern is actually read, and subject factors, because it is the students themselves who are doing the reading. Nevertheless, several problems arise from following such a procedure. First, there is strong evidence that tests which teachers

write are not reliable measures of content difficulty alone. Long ago, Lorge (1949) criticized direct testing because it was uncertain whether the actual difficulty level of the passage was reflected in the difficulty of the language used in the test items themselves. For example, a difficult selection may appear easier if the test questions the student must answer are simply phrased. Similarly, an easy passage may appear difficult if the test questions are couched in more difficult language than the language of the passage itself. Such factors, of course, are determined by individual teachers. Bormuth (1975) says:

> Test writers influence the difficulty of tests: two writers making a test over a single passage could produce tests of quite different difficulty, the one writer's test eliciting mostly low scores and the other's mostly high scores (p. 63).

Although directly testing students on passages they have read seems more valid than readability formulas as a measure of reading difficulty, such tests are nevertheless unreliable and suspect because the test items themselves also represent a reading task, the difficulty of which is determined by individual teachers and not by the reading selection itself. Because of this, Bormuth (1967b) indicates that "it is uncertain whether it is the difficulty of the passage or the difficulty of the (test) items that is measured by this procedure" (p. 5).

A second, more subtle, problem associated with directly testing the student's comprehension of a reading selection involves the manner in which the test items are

selected for inclusion in the test. Bormuth (1970) states the following:

> The traditional item writer seldom, if ever, sets out to write every possible item of a given type in making a test over a program. Rather, he avoids writing items he considers trivial, too complex, too simple, too wordy, or otherwise undesirable. Similarly, he will often simply decide that he has produced as many items as he needs. His decisions on these matters are seldom explicit and are almost never rationalized. Thus, as the test writer generates items by traditional methods, he is implicitly designing the test, but doing so in a manner that is not open to inspection and rational review (p. 13).

Because any set of questions chosen to test how well a student has comprehended a reading selection is nothing more than a sample of all the possible test questions that could be written over the selection, it is extremely important that the sample of test questions be an unbiased one. Unfortunately, however, because the manner in which questions are selected is so unrationalized and non-operational, there is no way to be certain that the final test is either biased or unbiased. There therefore arises another source of uncontrollable error variance that can have marked effects on the outcome of the test.

A third problem in directly testing comprehension involves the selection of acceptable criterion scores. According to Bormuth (1969), reputable reading authorities such as Betts (1946), Bond and Tinker (1967), and Harris (1962) claim that reading materials are suitable for unsupervised study if a student can score at least 90 percent on a test over their content. Materials are

regarded as suitable for supervised study if the student can score 75 percent. Bormuth further indicates that a search of the literature has revealed that those two criterion scores (75 percent and 90 percent) originally could be traced back to Thorndike (1917), who apparently acquired them from teachers who had begun using them as a result of oral tradition. Interestingly, however, is Bormuth's (1975) observation that none of these reading experts has ever offered any evidence, reasoning, or rational support for selecting these particular scores. In fact, Bormuth (1967b) states that traditionally the selection of a criterion score has been viewed "as a matter to be left to personal preference or arbitrary choice rather than as a matter for rational decision based, at least in part, on empirical data" (p. 22). Bormuth (1975) further suggests:

> We are left to wonder if the criterion scores sprang from an arbitrary decision that became standard practice simply because of the strength of precedence. The length of the English unit of measure, the foot, took on its value because the king who standardized that measure happened to have a foot of that length. The chief effect of this accident of history was to saddle us with an amazingly inconvenient measurement system (p. 64).

From all of the foregoing, it should be readily apparent that there is no empirical evidence of any sort suggesting that a criterion score of either 75 percent or 90 percent makes a passage ideal or, for that matter, even acceptable for instructional use.

The Cloze Procedure, A Third Possibility

Recognizing that both readability formulas and direct tests of reading comprehension possess inherent weaknesses that can bias and therefore invalidate findings, where next must one look in seeking a way to objectively assess the difficulty level of reading selections?

A procedure that has demonstrated great promise for assessing the difficulty of both native and foreign language reading selections is the cloze test. Introduced by Taylor (1953), who adopted the word "cloze" from the Gestalt psychologists' concept of closure, the cloze procedure is a way of constructing tests by mechanically and systematically deleting the words in a reading selection and replacing them with an underlined blank of standard length. Students being administered the tests are then expected to guess what word has been taken out of each blank and write it in that space. There is a variety of ways to select the words to be deleted. According to Bormuth (1973-1974), "what distinguishes a cloze test from an ordinary deletion test, though, is the fact that in a cloze test the words may be selected for deletion only by a completely replicable set of rules" (p. 38).

Upon close examination, one fact becomes immediately apparent regarding the cloze test: Because it is a technique that measures how well a reader has understood what he has read, it may be used both as an index of a passage's readability, i.e., its reading difficulty,

and as an index of a student's ability to comprehend the passage. This characteristic is important because it indicates that both material and subject factors are taken into consideration, which, of course, is essential in order for any reading measurement test to be valid and reliable. According to Anderson (1976):

> The significance of cloze procedure as a measure of readability is that the method appears to take into account both material and subject factors of readability. There is a sound theoretical basis also for cloze procedure as a measure of reading comprehension for cloze scores index the correspondence of language habits of writer to those of reader and thus the construct, reading comprehension, is given operational meaning (p. 7).

The cloze test appears to have the necessary qualifications for adequately dealing with both material and subject factors in assessing reading difficulty. In addition, however, unlike direct tests, which are traditionally influenced by the various biases of the individual test writer who, in generating questions, can make the reading selection appear either more difficult or easier than it actually is and thereby reduce the validity of the findings, cloze tests do not inject irrelevant sources of variance into the process because the manner in which the cloze test is generated is operationally defined in advance. Therefore, the test reflects only the actual difficulty level of the reading selection under consideration. For obvious reasons, such a procedure yields far more valid results than would be had by using either a traditional readability formula

or a direct test of reading comprehension.

A Problem with the Cloze Test

Despite the fact that the cloze procedure deals with both material and subject factors and is generated in such a manner that error variance is basically eliminated, there nevertheless exists a significant problem.

Bormuth (1967b) indicates that "the difficulty of a text should be reported in terms that make clear how appropriate the text is for a given individual or group" (p. 21). In determining how appropriate the reading selection is, one simply reports how many individuals are able to score at or above some predetermined criterion level. To do this, however, requires that cloze tests have a criterion score that represents an acceptable level of understanding.

The most obvious method of establishing a criterion score for the cloze test is to adopt a score traditionally used and then determine what cloze score is comparable with that criterion score. Reputable reading authorities have suggested that reading selections are suitable for supervised study if students can score 75 percent, and selections are appropriate for unsupervised study if a score of 90 percent can be achieved. Unfortunately, as has been pointed out, however, no one has ever offered any evidence supporting the validity of such scores, and it is therefore not appropriate to use them in establishing objective cloze criterion scores.

An Alternative Cloze Criterion Score

Because traditional scores have been established using subjective rather than objective standards and are therefore most suspect, where can one turn in order to determine meaningful cloze criterion scores? In carefully analyzing the problem, it is immediately apparent that a most important first step is to determine what is actually desired from reading. After making such a determination, the task, simply stated, should be nothing more than establishing the relationship between what is wanted from reading and the scores achieved by administering the cloze test.

While there are numerous desirable results from reading, both cognitive and affective, one of the most desirable is that the reader actually gain some factual information as a result of having read a given selection. If, as a result of reading, information is gained by the reader, it should be readily apparent that at least the following significant educational aims can be achieved: 1) Overall curricular objectives can be better met; 2) Students can have a much greater opportunity to be educationally rewarded and satisfied; 3) At a minimum, chances for student academic survival can be greatly enhanced. In establishing meaningful cloze criterion scores, therefore, it would appear most appropriate simply to determine the relationship between actual information gained and cloze scores. More specifically, it would seem most

valuable if one could determine which individual score on cloze tests, in general, reflects that point at which readers of German actually begin gaining substantial factual information. By knowing such a score, of course, foreign language teachers could make objective, meaningful statements concerning the appropriateness, with regard to difficulty, of any reading selection.

The Best Means for Measuring Information Gain

How does one proceed in establishing which cloze score reflects that point at which readers begin gleaning substantial information? As an initial step, one must determine how best to measure information gain. According to Bormuth (1967b), the ultimate test of assessing how much information has been gained is to subject the reader to the environment referred to in the reading selection and see whether he or she behaves appropriately. Such a procedure, however, is at best inconvenient, and in most cases impossible, because it would involve recreating events such as calamities and catastrophies simply to test a reader's knowledge of the events or situations. As alternatives, therefore, educators have resorted to procedures such as having students write essay examinations, give oral reports, or answer objective questions, either completion-type or multiple-choice, over the assigned reading material.

Although all three of the above-mentioned testing procedures are used in education to assess how much

information the reader has gained, which is most appropriate to be used in comparing information gain with cloze scores and thereby establishing objective and rational cloze criterion scores? In answering this question, the following should be considered:

Stated most simplistically, any reading selection can be viewed as nothing more than a list of content-related sentences. Unlike essay examinations or oral reports, conventional objective tests, in their simplest form, are constructed by drawing a predetermined number of selected sentences from the list and transforming them into questions to be answered either with a short sentence (completion-type) or by selecting from several choices the correct answer (multiple-choice). In this situation, transforming the sentence into a question has the effect of deleting that portion of the sentence that becomes the correct response to the question. In answering the question, therefore, one simply supplies that segment of the sentence that was deleted by the transformation. As a most elementary example, if "the boy" in the sentence "The boy threw the rock into the pond." is considered to be content significant and is therefore selected from a passage to serve as a test question, the sentence is transformed into a question by simply replacing "the boy" with the interrogative "who," which results in the formation of the question "Who threw the rock into the pond?". In a completion-type objective test, the answer

would be "the boy"; in a multiple-choice test, "the boy" would be included with two or three distractors. In either case, however, the answer would be provided by supplying the word that had been deleted as a result of transforming the sentence into a question.

From the foregoing, it is readily apparent that both the cloze test and the conventional objective test are constructed by generally like manipulations and therefore demonstrate at least a formal similarity. Because of this similarity, the objective test is the test type best suited for establishing the relationship between information gain and cloze scores and thereby determining meaningful cloze criterion scores.

An Objective Test Weakness

In determining the relationship between the cloze test and a conventional objective test in order to establish cloze criterion scores, one major problem becomes immediately apparent. Although the two types of tests demonstrate formal similarity, a chief distinction between the two is found in the method employed to select the items for use in the test. As has already been pointed out, cloze tests are constructed by mechanically and systematically deleting the words in a reading selection according to a completely replicable set of predetermined rules. Conventional objective tests, on the other hand, are generated in a very subjective manner, being greatly influenced by the biases of the individual test writer,

whose rationale for constructing the various test items is generally based upon personal feelings about what he or she calls "important knowledge." In assessing the relationship between the cloze test and a conventional objective test in order to establish a rational cloze criterion score based on information gain, it is extremely important that not just the cloze test, but that both tests be completely objective, unbiased, and free from all subjective considerations. Only when this is accomplished may the test items then be regarded as the property of the reading selection itself with no other factors systematically influencing the tests, and only then can one be sure that all the findings will be unbiased and therefore valid and generalizable.

Objective

Despite the fact that reading ultimately determines successful foreign language learning more than any other of the four language skills, assessing the difficulty level of German reading selections in order to choose appropriate reading passages has proven to be a rather arbitrary and unexplicit process. Of the three procedures presently designed to measure reading difficulty (readability formulas, direct tests of reading comprehension, and cloze tests), only the cloze procedure has been shown to have definite promise because it alone deals with both material and subject factors, is unbiased by the test writer, and can be easily operationally replicated.

Nevertheless, failure by reading experts to establish objective guidelines for interpreting cloze scores has largely invalidated the cloze procedure as a tool for accurately determining the difficulty level of German prose reading selections.

The overall goal of this investigation, therefore, is to establish empirically a meaningful and valid cloze criterion score that 1) objectively reflects passage difficulty and 2) will be valid for any German prose reading selection to which the cloze test can be applied. Because actually acquiring information is one of the most desirable outcomes of reading, the cloze criterion score to be established will be based on information gain. In this regard, the objective will be to determine that score on the cloze test at which nonnative readers of German begin to gain substantial information as a result of reading. Expanding on extensive cloze research carried out by Bormuth (1971) in English, this investigation will seek to establish a cloze criterion score based on information gain by examining and establishing the relationship between the cloze test and the objective test, an accepted measure of information gain that has a demonstrated formal similarity to the cloze test. More specifically, the cloze scores a carefully chosen sample of college German students achieve over a selected piece of German prose will be empirically compared and analyzed with the objective test scores carefully matched counterparts achieve after

reading the same selected piece of German prose.

In order to control subjectivity and individual test writer biases and thereby ensure that the cloze criterion score established as a result of this research is both valid and generalizable to any German prose reading passage to which the cloze test might be applied, it is essential that both the cloze test and the objective test be created following operationally replicable procedures. Although cloze tests are, by nature, replicable, such is not the case with objective tests. Therefore, there arises the requirement for generating an objective test-writing algorithm to ensure the objectivity and replicability of the conventional objective test. Using ideas set forth primarily by Bormuth (1970) and Finn (1973), but altered to apply to a German prose reading selection, such an algorithm will be generated as part of this research project. Details of the rules used to generate the algorithm will be discussed in Chapter III. In general, however, the algorithm is designed to generate unbiased and operationally replicable, short-answer, completion-type objective test questions for use in this research project. The completion-type objective test was chosen over the multiple-choice type based on research reported by Bormuth (1971), which indicated that greater conformity to the actual difficulty of the reading selection itself could be achieved by using short-answer, completion-type test questions.

Based on Bormuth's (1962 and 1971) initial findings in English, it is anticipated that the relationship between the two kinds of tests (cloze and objective) in this research will likely be nonlinear. In discussing the cause for this nonlinearity, Bormuth (1971) indicates that according to his findings, the cloze test, although similar to the conventional objective test, appears to be both more difficult as well as capable of discriminating over a broader range of reading abilities than does the objective test. For example, on an easy reading selection, objective test score distributions frequently show ceiling effects due to large numbers of students obtaining perfect or very high scores. On the other hand, on difficult passages, objective test score distributions often exhibit flooring effects as a result of some students scoring at or near zero. While cloze scores also exhibit flooring effects, they seem to do so to a lesser degree, and rarely, if ever, do they show ceiling effects. According to Bormuth (1971), these facts suggest that there is a broader range of skills involved in reading than is normally identified and measured by objective tests. This also suggests that the cloze test is capable of measuring both those skills that are so simple and easy that they fall below the lower limits of the objective test as well as those skills that are so complex and difficult that they fall above the upper limits of the objective test. Because of this greater discriminating capability of the cloze

test, it can therefore theoretically be anticipated that in the lower range of cloze scores, information gain scores, as reflected by the objective test, will show little increase as cloze scores increase. This, because some of the items in the cloze test are easier than the easiest of those in the objective test. Therefore, the objective test will theoretically fail to discriminate among students scoring in the lower ranges of the cloze tests, which will cause the slope of any plotted line to stand at or near zero. When scores fall into the higher ranges on the cloze test, however, increases in the objective scores will also be observable because the two tests will both be discriminating in that range. It is at this point that the slope of the plotted line will approach linearity.

As might be expected, that point at which both the cloze scores and the objective test scores begin significantly rising together can be viewed as that point at which the student begins to gain substantial information as a result of reading as reflected both by cloze and objective test scores. This point, of course, can be used justifiably as a rational and objective cloze criterion score because it reflects that level on the cloze test that students must achieve in order to begin gaining substantial information as a result of reading.

Summary and Statement of Research Questions

In achieving general academic success, reading has been identified as one of the most important skills. Indeed, no comprehensive foreign language program is conceivable without ultimately involving large quantities of reading. A most important factor in successful reading, however, is that the student be given selections whose difficulty level is within his or her range of reading comprehension. Little in education is more frustrating than struggling with reading material far too advanced for one's capabilities.

At the present time, the assessment of foreign language reading difficulty, if objectively assessed at all, is determined generally either by applying some type of readability formula to the passage, or by directly testing the reader on the reading selection in question. Recently, however, the validity of such methods for assessing reading difficulty has been so seriously challenged that the value of such procedures as reading assessment tools is seriously questioned. As a result, therefore, educators have been looking for a different, more valid means by which to assess reading difficulty.

A procedure that has demonstrated great promise because of its objectivity, replicability, and ability to take into account both material and subject factors affecting readability is the cloze test. Nevertheless, a prime hurdle in effectively using the cloze test as

a means of determining the difficulty of reading selections has centered around the establishment of meaningful and valid cloze criterion scores. The obvious solution of establishing which cloze score equates to the 75 percent level of a conventional objective test, a standard established and accepted by most reading authorities, has been rejected because such percentage scores have proven to be arbitrary, subjective, and highly questionable.

As a more meaningful and objective alternative to the above, it has been decided to explore the possibility of comparing the relationship of the cloze test to what is actually most desired from reading, i.e., the gleaning of factual information. In making this comparison, German prose cloze-test scores obtained from carefully matched, college-level, nonnative speakers of German are to be compared with scores that other matched, college-level, nonnative speakers of German achieve on a specially prepared, operationally replicable, short-answer completion test constructed over the same piece of German prose. As a result of this comparison, it is hoped to be able to establish an objective, meaningful cloze criterion score that can be applied to any German prose reading selection, and that does, in fact, reflect that cloze level at which nonnative readers of German prose begin to gain substantial information as a result of reading.

The purpose of this research, therefore, is to determine the answers to the following questions:

1. Is there a relationship between the cloze test and information gain, as measured by a specially constructed, short-answer completion test, when both test types are prepared from a single German prose reading selection, selectively administered to a carefully chosen group of college-level nonnative speakers of German, and then statistically analyzed?

2. If there is a relationship between the two test types, will it be nonlinear, as Bormuth's work in English suggests, when applied to nonnative speakers of German?

3. If the relationship between the two test types is nonlinear, can a rational, objective cloze criterion score that is valid for any German prose reading selection be identified, which reflects that point on the cloze test at which nonnative readers of German begin gaining substantial information as a result of reading German prose?

Overview

The balance of this research report is divided into four chapters followed by the appendices and bibliography. In Chapter II, relevant literature is reviewed. Chapter III describes in detail the instruments and procedures employed in this research. In Chapter IV, the statistical analysis and findings are presented and discussed. Chapter V provides a summary of the study including a discussion of the limitations, conclusions, implications, and specific recommendations for additional research.

CHAPTER II

REVIEW OF LITERATURE

Introduction

Objective

The purpose of this research project is to establish an objective, unbiased, and meaningful cloze test criterion score that can be used by foreign language teachers as a basis for choosing appropriately difficult German prose reading selections. To accomplish this goal, this research study has been designed to investigate the relationship between the cloze test and information gain, as measured by an operationally replicable, short-answer, completion-type objective test. The objective is to establish at which point on the cloze test do nonnative readers of German actually begin gaining substantial factual information. It is therefore appropriate to address the following areas of concentration: 1) The cloze test as a measuring instrument; 2) The cloze test and foreign language research; 3) the cloze test and criterion scores.

The Cloze Test as a Measuring Instrument

Cloze Development

Since its introduction by Wilson Taylor in 1953, the cloze test has stimulated considerable interest and research activity. Originally intended by Taylor to be a "new and better" way of assessing the readability of printed English prose, its utility in a variety of other research areas quickly became apparent. After two years of experimentation, Taylor (1956) indicated that "the cloze method appears to possess a very large number of potential research uses, both practical and theoretical" (p. 42).

In 1957 Rankin completed his doctoral dissertation on the cloze procedure and in 1959 wrote the first review of research on studies using the cloze method. Although Rankin's review referred to only a handful of studies using the cloze test, the procedure subsequently began to grow rapidly in popularity as a research tool. In the 1960s much research and many doctoral dissertations covering a wide variety of cloze-related topics were completed, and in 1968 the National Reading Conference conducted a cloze symposium. Since that time the cloze test has received ever-increasing attention from researchers.

Cloze Uses

At the present, research involving the cloze procedure is generally divided into three broad categories: 1) research investigating cloze methodological considerations

in order to improve the validity and reliability of findings; 2) research in which the cloze test is examined as a reliable and valid measuring device; and 3) research in which the cloze test is examined for its utility as a teaching technique. To date the majority of research has been directed toward examining the cloze procedure in its role as a device for measurement.

As a measuring instrument, the cloze test has been used widely to study the reading communication process, both from the standpoint of the message and its characteristics (readability) as well as from the standpoint of the message consumer (comprehension). In this regard, it has been used: 1) to assess and predict readability at all levels, in a variety of ways, and in various languages; 2) to probe the difficulties readers encounter in prose; 3) to isolate precise linguistic variables associated with readability; 4) to measure both general and specific reading comprehension; 5) to uncover certain personality and physical disorders; 6) to evaluate aural comprehension; and 7) to assess verbal apptitudes. As can be seen from only this partial list of cloze applications as a measurement tool, a review of all pertinent literature related to the various facets of just measurement is far beyond the scope of this project. In fact, in discussing the widespread use of the cloze test, Oller (1973) says that it has "been constructed in so many ways and for so many purposes that an overview of

the entire scope of literature on the subject is challenging to the imagination not to mention the memory" (p. 105).

Pertinent Reviews

Because so much cloze research has been conducted during the past 15-20 years, in this work it is necessary only to point out that many valuable works devoted primarily to reviewing cloze literature have been published since Rankin's (1959) initial review. Those most pertinent and comprehensive reviews examined in conjunction with this research project include the following: 1) The cloze Procedure--A Survey of Research, by Rankin (1965); 2) Cloze Readability Procedure, by Bormuth (1967a); 3) A Taxonomy of Cloze Research, Part I: Readability and Reading Comprehension, by Potter (1968); 4) The Cloze Procedure: A Conspectus, by Bickley, Ellington, and Bickley (1970); 5) The Cloze Procedure: A Survey of the Research, by Jongsma (1971); 6) A Review of Literature Related to the Cloze Procedure, by Fram (1972); 7) A Comprehensive Bibliography of the Cloze Procedure, by Boyce (1974); 8) Historical Overview of the Cloze Procedure, by Walter (1974); 9) Research with Cloze Procedure in Measuring the Proficiency of Non-Native Speakers of English: An Annotated Bibliography, by Oller (1975); 10) Psycholinguistic Experiments in Foreign Language Testing, by Anderson (1976).

The Cloze Test and Foreign Languages

Since the inception of the cloze procedure in 1953, almost all research has been conducted in English using native speakers of that language. Oller (1973) confirms this statement, indicating that until very recently nearly all cloze investigations used native speakers of the language involved, and "that in practically all of the published studies the language used was English" (p. 107).

Cloze studies involving foreign languages and non-native speakers have been carried out on a sporadic and rather infrequent basis. In reviewing foreign language cloze research, Oller (1975) says that although many language teachers and bilingual educators have begun to investigate the cloze procedure, "it is probably true, nonetheless, that the technique has been vastly underexploited to date" (p. 1). Although some work has been done examining primarily methodological considerations (Oller, 1972b, for example), most foreign language cloze research has had as its ultimate objective the examination of the cloze procedure's reliability and validity as a measuring device. In this regard, two areas have been given considerable attention: 1) measuring global foreign language proficiency, which is of less importance to this study, and 2) measuring foreign language readability and/or reading comprehension.

Global Foreign Language Proficiency

At the present time, considerable emphasis is being placed on the cloze procedure as a tool for assessing foreign language proficiency. First to address this topic were Carroll, Carton, and Wilds (1959), who according to Rankin (1965), concluded that the cloze test could be used to measure group differences in foreign language competence (German and French) but was not suitable for assessing individual differences because "the tests tended to have low reliability and were heavily affected by various sources of extraneous cognitive factors which are independent of competence in a foreign language" (p. 140). The validity of these findings is suspect, however, for at least three reasons: 1) In preparing the cloze test, the random deletion procedure was not used. 2) According to Oller (1975), the study utilized "a very small sample for many of its generalizations" (p. 20). 3) Instead of using a proficiency measure for validating, the Modern Language Aptitude Test (MLAT) was used.

Since Carroll, Carton and Wild's research in 1959, assessing foreign language proficiency, especially in English-as-a-second-language (ESL) situations, has become very popular. In this regard, Jones (1977) indicates that Oller (1972a, 1975, 1976a, 1976b) has conducted extensive research and that the cloze procedure "has been shown to be an efficient, valid and reliable method of measuring language proficiency" (p. 257). Nevertheless,

although cloze research in ESL proficiency has been widely investigated, especially by Oller, cloze investigations designed to measure the foreign language proficiency of native speakers of English have been almost nonexistent. Since Carroll, Carton, and Wild's original study in 1959, this author was able to uncover only two other published cloze research projects (Teitelbaum, Edwards, and Hudson, 1975; and Briere et al, 1977) and two recently completed, but as yet unpublished cloze investigations (Clausing and Lange, 1977; and Hanzeli, 1977) designed to measure the foreign language proficiency of native speakers of English.

Foreign Language Readability

In addition to using the cloze test to measure foreign language proficiency, limited interest has also been expressed in cloze measurement of foreign language readability and/or reading comprehension. In reviewing foreign language readability research, only four studies were detected:

1) Taylor (1954) was the first researcher to apply the cloze technique to a foreign language. Using Korean, because it was a language syntactically and symbologically totally unrelated to English, Taylor mailed three cloze tests to a sample of native Korean college students located throughout the United States. Prior to mailing, each passage was subjectively judged as easy, medium, or difficult. Although only about 21 percent of the

passages were returned, analysis of the cloze scores ranked the three passages in the same order as the initial "a priori," subjective rankings had placed them.

2) Three years later, Shiba (1957) found the cloze procedure to be applicable to Japanese. Using eight Japanese passages that had been ranked by three judges, Shiba reported a correlation of .83 between cloze scores achieved by native Japanese and subjective judges' ratings.

3) Anderson (1976) reports that according to research carried out by Klare, Sinaiko, and Stolurow (1971), the cloze procedure has proven to be a valid and reliable tool for discriminating between good and bad translations of English into Vietnamese.

4) In the only known cloze readability study in which nonnative speakers of the foreign language were involved, Kerr (1968), utilizing nonnative speakers using French reading materials, replicated Taylor's initial studies. In Kerr's investigation three passages were initially subjectively ranked by experienced French teachers. Three different forms of the cloze were then prepared over each passage, in which either every fifth, every tenth, or random ten percent deletions were made. The three versions were then administered to Australian students studying French. Despite the different ways in which the cloze tests were prepared, there was a perfect correspondence between teacher ratings and cloze score.

Foreign Language Reading Comprehension

Only one study devoted solely to foreign language reading comprehension was detected. Using foreign students studying at the University of Florida, Friedman (1964) found that cloze scores on 20 English reading selections correlated significantly with the Metropolitan Achievement Test.

The most comprehensive foreign language cloze research utilizing the cloze procedure as a measure of both readability and reading comprehension was reported by the Australian Anderson (1976). As part of an ongoing project, Anderson conducted a series of ten experiments designed to evaluate various methodological considerations as well as the validity and reliability of the cloze test as a measure of readability and reading comprehension. Subjects for the first eight experiments were second-language learners from Papua, New Guinea and Fiji, and all reading materials were in English. For the last two experiments, subjects were drawn from Malay- and Chinese-medium schools in Singapore, and all reading materials were in Malay and Chinese respectively. In summarizing the overall findings of the research, Anderson (1976) says the following:

> To conclude, evidence from the ten experiments has indicated the usefulness of cloze procedure as a method of measuring reading difficulty and of testing reading comprehension with subjects for whom English is a second or foreign language, and to a less extent, with native speakers of Malay and Chinese. The wider application of cloze procedure

> to cross-cultural and foreign language testing is a development on the horizon which offers exciting prospects (p. 132).

As was the case in cloze research intended to assess foreign language proficiency, few cloze studies designed to measure foreign language readability or reading comprehension have involved native speakers of English using a foreign language. Except for Kerr's (1968) research, no other evidence of such investigations was detected.

The Cloze Test and Criterion Scores

Although considerable research has been carried out to assess the value of the cloze procedure as a tool for measuring reading comprehension, especially in English, the problem of objectively interpreting cloze test results has continually perplexed investigators. So serious has the barrier been that Walter (1974) indicates: "The problem of how to interpret cloze test scores has plagued researchers more than any other facet of the cloze procedure" (p. 23).

Early Approaches

In initial cloze research, investigators used raw scores to interpret results. This method is of little value, however, because cloze tests vary in length. Therefore, at best, the raw score can only have meaning for each individual test.

In an attempt to circumvent the problem created by using raw scores, many researchers began converting raw scores to percentages. Although preferable to the raw

score because it allows comparisons between cloze tests of different lengths, the cloze percentage score is only slightly better than the raw score because it also does nothing more than reflect a rank ordering. For example, if student X scores 50 percent and student Y scores 40 percent on a cloze test intended to measure reading comprehension, one can only safely assume that student X performed better on the cloze test than did student Y. What is not known, however, is how precisely each student comprehended the selection.

Finding Equivalent Scores

Because neither the raw score nor the percentage score was a useful and meaningful cloze criterion, investigators began examining how the dilema could be resolved. A logical and direct approach to the problem seemed to be to adopt a cloze criterion score that was equivalent to a performance level traditionally accepted. For years reading authorities have accepted 75 percent and 90 percent as adequate supervised and unsupervised performance levels in assessing reading comprehension. Because of this broad acceptance, Bormuth (1967c) therefore decided to determine the cloze scores that were equivalent to the traditional 75 percent and 90 percent reading comprehension levels.

Selecting 100 students enrolled in grades four and five, Bormuth used a regression prediction equation to match cloze test results with performance on multiple-

choice comprehension tests over the same nine passages. His findings indicated that a cloze score of 45 percent was comparable with a multiple-choice comprehension score of 75 percent and that a cloze score of 52 percent was comparable with a comprehension score of 90 percent.

Not completely satisfied with these results, however, the next year Bormuth (1968) determined to establish which cloze scores equated to the 75 percent and 90 percent criteria when measured with completion-type comprehension tests administered during an oral reading test. Taking 120 students enrolled in grades four, five, and six, Bormuth administered all four equivalent forms of the Gray Oral Reading Paragraphs, each form of which consists of 13 paragraphs graded according to difficulty. First, every subject was given a cloze test over two of the four possible paragraphs at each level of difficulty. Then, all were given an orally administered, completion-type comprehension test over each of the two remaining paragraphs immediately after having read each paragraph orally. The cloze scores equivalent to the criterion scores were established by finding the most difficult level at which a student was able to obtain first a score of 75 percent and then a score of 90 percent on the two sets of comprehension tests. When this level was determined, the two cloze test scores achieved over the other two comparable selections were averaged to determine the cloze criterion score. Results of Bormuth's findings in this

study reflected that a cloze score of 44 percent was comparable with an orally administered, completion-type comprehension test score of 75 percent and that a cloze score of 57 percent was equivalent to the 90 percent level on that type of test.

In explaining the difference between the two cloze criterion scores established with the multiple-choice tests as compared with those found with the completion test (see Table 1), Bormuth (1969) says the 90 percent criterion score on the multiple-choice test was most likely depressed by a ceiling effect. Therefore, the score obtained in the 1968 study "seems to be a better estimate of the cloze score comparable to the 90 per cent criterion" (p. 62).

Recognizing the importance of establishing meaningful and objective cloze criterion scores, researchers have conducted two investigations to replicate Bormuth's 1967 study. Rankin and Culhane (1969) selected 105 students in grade five and compared their cloze test results with multiple-choice tests over the same five reading selections. Results indicated that a cloze score of 41 percent was comparable with a multiple-choice comprehension score of 75 percent, while a cloze score of 61 percent equated to the 90 percent comprehension score.

In the only study in which nonnative speakers of the language were involved to establish a meaningful cloze criterion score, Anderson (1976) selected 284 elementary-

TABLE 1

RESEARCH FINDINGS BY BORMUTH DESIGNED TO ESTABLISH CLOZE CRITERION SCORES EQUIVALENT TO TRADITIONALLY ACCEPTED PERFORMANCE LEVELS

Traditional Scores	Comparable Cloze Scores	
	1967 Study (Multiple Choice)	1968 Study (Completion)
75	45	44
90	52	57

school students from three separate schools in Papua, New Guinea. Using English passages ordered according to difficulty, both a cloze test and a multiple-choice comprehension test were constructed over each passage. Students were first administered the cloze tests and then one week later were given the multiple-choice tests. As a result of regression analysis, Anderson reported that a cloze score of 44 percent equated to a multiple-choice comprehension score of 75 percent. A cloze score of 53 percent was equivalent to the 90 percent comprehension score (see Table 2).

TABLE 2

CLOZE TEST CRITERION SCORES EQUIVALENT TO
TRADITIONALLY ACCEPTED PERFORMANCE LEVELS

Traditional Scores	Comparable Cloze Scores			
	Bormuth		Rankin and Culhane	Anderson
	1967 Study (Multiple Choice)	1968 Study (Completion)	1969 Study (Multiple Choice)	1976 Study (Multiple Choice)
75	45	44	41	44
90	52	57	61	53

Two Variations

In a related attempt to establish a meaningful cloze criterion score, Alexander (1968) gave 365 elementary-level students cloze tests and the Gates-MacGinitie Reading Tests--Survey D. A statistical analysis that compared the results revealed that a cloze score any lower than 47 percent resulted in student frustration, while a cloze score of 61 percent or above suggested the student could read the material independently. These two percentages (47 and 61), of course, equate to Bormuth's 75 and 90 percent levels.

In a slightly different and somewhat questionable approach, Ransom (1968) arbitrarily established the cloze score of 50 percent to represent an independent reading level, 30 to 50 percent to represent and instructional level, and below 20 percent to represent a frustration level. She then correlated her predetermined cloze scores with the results of an Informal Reading Inventory. Because correlations were high, she assumed her comparable cloze scores were valid for estimating students' reading levels from grades two through six.

Problems with Equivalent Scores

Although all of the research studies mentioned above are significant steps toward establishing meaningful and objective cloze criterion scores, there are nevertheless serious methodological weaknesses that make their findings suspect. Disregarding individual weaknesses of design,

such as Ransom (1968) arbitrarily selecting her cloze criterion scores, the following should be pointed out: In conducting their investigations, all of these researchers (Bormuth, 1967, 1968; Alexander, 1968; Ransom, 1968; Rankin and Culhane, 1969; and Anderson, 1976) analyzed their cloze test results with some form of a direct test of reading comprehension. As was pointed out in Chapter I, direct tests of reading comprehension have at least three basic weaknesses:

1) There is strong evidence that such tests are not reliable measures of content difficulty alone because test writers tend to write questions in such a fashion that the language used is either easier or more difficult than the language in the actual reading selection. If the test reflects anything other than the actual difficulty of the reading selection, the cloze criterion scores that are subsequently established will be proportionately invalid.

2) The manner in which test items are first generated and then selected for inclusion into the reading comprehension test is suspect. Instead of systematically generating a population of possible test questions in an operationally replicable manner that is open to investigation and then sampling from that population to determine the specific questions, test questions generally are created according to biases and preferences of each individual test writer. Again, the result is a distortion of the findings. In addressing this very problem, Bormuth (1969) says of his

1967 and 1968 studies that they "share the problem that the population of traditional comprehension test items used were not rigorously defined and then sampled to compose the tests" (pp. 62-63).

3) There is no evidence anywhere in the literature that supports the supposition that the criterion scores of 75 percent and 90 percent are ideal. No reading expert has ever offered any reasoning or rational support for such scores. Instead, these two figures have arisen as a result of tradition--certainly a poor criterion for selecting meaningful cloze criterion scores.

Considering the basic weaknesses associated with direct tests of reading comprehension used in the studies discussed above, it is evident that any resultant cloze criterion scores must be viewed, at best, as very questionable. In discussing three of the above studies, Jongsma (1971) points out that they are honest attempts at building a framework to help in making value judgements about cloze test results. There are limitations, however:

> They all are based on the notion of independent, instructional and frustrational reading levels and the commonly used percentages for defining such levels. The research that the definitions of these levels is based upon is far from convincing or conclusive. Any attempt to interpret cloze test results on the basis of such rationale must be treated cautiously (p. 20).

Some Groundwork Involving Information Gain

Accepting the fact that the validity of cloze criterion scores is suspect when established by finding what cloze score equates to 75 percent and 90 percent on traditionally accepted reading comprehension tests, researchers interested in determining more objective cloze criterion scores began investigating other avenues.

The groundwork for a more objective approach to establishing cloze criterion scores was laid by Coleman and Miller (1968) whose research objective was to determine what level of cloze passage difficulty resulted in the greatest amount of information gain. Using undergraduate college students, they measured information gain by typing a reading passage on a transparency and then covering the words with strips of tape. When this was projected on an overhead, the students were asked to guess and write down the first word. This word was then exposed and the students were then asked to guess and write down the next word. After all words in the passage had been guessed in this manner, the tape was replaced over the passage and the procedure was repeated. The difference between scores on the two trials was taken as a measure of information gain. Passage difficulty was established on a matched group of subjects who were administered cloze tests. Of interest is the fact that results indicated that maximum information gain occurred on passages having difficulties of close to 44 percent--the cloze score Bormuth (1967c, 1968)

found to be equivalent to the traditional 75 percent criterion.

Although Coleman and Miller's research could be criticized (MacGinitie, 1961; Walter, 1974, for example) because the information gained may be overly influenced by rote memorization, Bormuth (1967a) says that "it seems clear that this study (Coleman and Miller, 1968) demonstrated how a rational approach can be made to the establishment of criterion scores" (p. 19).

Bormuth's Approach

Expanding on Coleman and Miller's (1968) belief that meaningful cloze criterion scores could be rationally established by investigating the relationship of the cloze test to information gain, Bormuth (1971) conducted a series of five studies ultimately designed to investigate the relationship of cloze scores to not only the cognitive variable of information gain, but to five other variables as well (rate of reading, willingness to study, and preferences for the subject matter, style, and level of difficulty).

In Bormuth's research, the intent was not to establish at which point on the cloze continuum comprehension began to occur. Instead, his objective was to determine at which point on the cloze test did the student achieve maximum comprehension. Although such an objective is slightly different from that of this study, the means by which these two objectives are accomplished are, in

this instance, generally the same. Because just the first three studies in Bormuth's series deal with objectively establishing the relationship between cloze scores and information gain, only they will be treated in this review.

Some Assumptions

In attempting to improve the objectivity of his research, Bormuth (1971) established three assumptions taken from psychometric theory, which he felt were essential in order to establish meaningful and objective cloze criterion scores. These he referred to as the instructional conformity, the behavioral consistency, and the regression identity assumptions.

The instructional conformity assumption asserts that in order for any comprehension test made from a reading selection to be valid, it must represent only the characteristics of the passage itself and not some other irrelevant considerations. This assumption, of course, addresses the problem of writing operationally replicable comprehension tests that have not been arbitrarily manipulated or tampered with as a result of the supposed "good judgement" of the test writer himself.

The second assumption, the behavioral consistency assumption, refers to the relationship among the behaviors measured by tests that are of the same type but that are made from different passages. This assumption asserts that tests of the same kind made from different reading passages must measure the same type of behavior on all

of the reading passages. Bormuth (1971) says that this can be accomplished "by making the same type of items for each passage" (p. 29)--something the cloze test does but most traditional reading comprehension tests do not do because they are constructed in such a subjective manner. Bormuth indicates, however, that traditional reading comprehension tests also can be made to comply with the behavioral consistency assumption by constructing them in an operationally replicable manner. In support of this, research carried out by Bormuth, Manning, Carr, and Pearson (1970) is cited, in which the researchers first defined large numbers of classes of comprehension items, constructed one item from each of a number of different paragraphs, and finally demonstrated that items generated by the same rules exhibited behavioral homogeneity despite the fact that they were written from different passages.

The final assumption, the regression identity assumption, asserts that a given type of test used to establish a performance criterion must exhibit "parametrically identical regressions on the independent variable, the cloze tests, regardless of the passage from which it is made or the student to whom it is given" (Bormuth, 1970, p. 30). This is most significant because if the regressions for some type of comprehension test differ from passage to passage, the cloze criterion scores will also differ, and it will therefore be impossible to establish cloze criterion scores for anything broader than

each individual reading selection. By the same token, if the resultant regressions differ for various groups of readers, different criteria will have to be established for each individual group.

Study One

In his initial study, Bormuth (1971) selected 260 English-speaking students ranging from grade three through graduate school. In the first of three separate testing sessions, the students were administered a 52-item cloze test in order to form individual pairs of readers matched for reading comprehension ability. To match the students, Bormuth ranked them according to the size of their scores on the cloze test and then matched them in order, a pair at a time. After matching, one member of each pair was randomly designated member X and the other member Y.

Two short reading passages were then selected (469 words and 398 words) and designated passages A and B. From each of these two passages were constructed two types of tests, a four-alternative, multiple-choice comprehension test (passage A, 34 questions and passage B, 39 questions) and five different forms of a cloze test in which every fifth word was deleted. In constructing five separate forms of the cloze test, the first form had words 1, 6, 11, etc. deleted; the second form had words 2, 7, 12, etc. deleted, etc. until all five forms of the cloze test had been constructed over each of the two passages, A and B. Five different forms of the cloze test were used

on each passage to counterbalance any effects that might arise should the items in one form differ in difficulty from the average difficulty of the items in all five forms. To ensure this, the five forms were randomly assigned to students so that each form occurred equally often throughout the entire study.

At the second testing session, all of the students designated member X were given the multiple-choice test made from passage A (without having first read passage A) and one of the five randomly assigned forms of the cloze test over passage B. At the same time, the Y members were administered the multiple-choice test made from passage B (without having first read passage B) and one of the five randomly assigned forms of the cloze test made from passage A. The multiple-choice tests were given prior to allowing the students to read the passages in order to establish how much prior knowledge the students possessed about the passage.

In the final testing session, conducted about 10 days later, the X members were asked to read passage A and then retake the multiple-choice test over it, while the Y members were doing the same thing with passage B. The testing sequence followed during this study is shown in Table 3.

In preparing his data for analysis, Bormuth computed the cloze scores over the two passages by finding the percentage of cloze items that each student had answered correctly. Information gain scores were computed by first

TABLE 3

BORMUTH'S STUDY ONE TESTING SEQUENCE

Pair Member	Prereading Multiple-Choice Test	Cloze Test	Postreading Multiple-Choice Test
X	Passage A	Passage B	Passage A
Y	Passage B	Passage A	Passage B

correcting all multiple-choice scores for guessing and then adjusting and subtracting the prereading scores from the postreading scores. These information gain scores also were then expressed as percentage scores.

The actual analysis consisted of taking each pair's cloze and information gain score over each of the two reading passages and comparing them by means of regression analysis. Upon finding that the two resultant curves were very similar, Bormuth combined the two sets of data for each pair and reanalyzed his findings. Speaking very generally, his findings suggested that students who scored no more than 25 percent on the cloze test gained little, if any, information from the reading passages. Above that point, there was a sharp and steady increase in the amount of information gained. According to Bormuth, at least three things made the findings of this initial study somewhat cloudy and questionable:

1) The two multiple-choice tests used for the study, although very carefully and thoughtfully constructed, were not systematically generated in an operationally replicable manner and therefore probably reflected something other than the actual difficulty of the two reading passages. As a result, both Bormuth's assumption of instructional conformity and his assumption of behavioral consistency were violated and the findings became suspect.

2) Only two reading passages, both similar in difficulty, were used in the study. Because more passages

of varying difficulty might have significantly altered the shapes of the regressions, Bormuth could not definitively say that this research confirmed and validated his regression identity assumption--something absolutely essential in establishing meaningful cloze criterion scores.

3) Bormuth's analysis of his findings reflected the unexpected tendency for older students to gain slightly more information at each cloze level. This suggested, of course, that, at least for children who were in the process of developing their reading skills, a slightly different cloze criterion score might have to be established for each grade level.

Study Two

Recognizing the weaknesses inherent in Study One, but also seeing the great potential for taking such an approach, Bormuth (1971) conducted his second study, which was designed primarily to accomplish two broad objectives: 1) to provide greater controls for the three problems that had been identified in Study One, and 2) to consider the effects of the students' attitude toward the reading materials, a factor Bormuth felt might be of significance.

Because using matched pairs of students tested during several (three) different sessions had resulted in large losses of cases due to absences at one or more of the testing sessions in Study One, Bormuth designed Study Two so that all data could be gathered in a single

testing session. To do this, each student was given a cloze test over one passage and a preference-rating scale and pre- and postreading completion-type tests all made from a second passage of matched difficulty.

To conduct the study, Bormuth selected 1320 students enrolled in grades three through 12. He then took 44 short reading selections equally drawn in groups of 11 from the four different forms of the Gray Oral Reading Paragraphs (1963 edition) and arranged them according to difficulty from very easy to very hard so that there existed four separate, but equivalently difficult, reading passages at each of 11 difficulty levels. From each of these 44 reading passages he then constructed five different forms of a cloze test in which various combinations of every fifth word were deleted. For example, the first form had words 1, 6, 11, etc. deleted; the second form had words 2, 7, 12, etc. deleted, etc. until all five forms of the cloze test had been constructed over each of the 44 passages. In addition, a ten-item, completion-type comprehension test was constructed covering each of the 44 passages. The completion-type comprehension test was chosen over the multiple-choice type because of difficulty Bormuth had encountered in Study One in writing acceptable alternatives that were unbiased and reflected only the difficulty of the passage, and because Bormuth (1970) had developed a more objective completion-type question writing procedure based upon wh- rote questions constructed from wh- words

and phrases such as who, what, where, when, what kind of, how many, etc. Finally, a seven-point interest scale ranging from "dislike very much" to "like very much" was devised to measure the students' willingness to study the textbook that a passage supposedly represented.

Using all of the above materials, test booklets were then constructed, each of which contained various tests made from one of two passages of similar difficulty. Found on pages two through five (page one contained student identification data) in each test booklet were the following: a completion-type comprehension test over one of the two passages (p. 2); a cloze test made from the second passage (p. 3); the passage from which the first completion test on page two had been made and the rating scale (p. 4); the same completion-type comprehension test the student had taken on page two prior to reading the passage (p. 5).

In assembling these test booklets, a counterbalanced design was used in which the four passages at each of the 11 different difficulty levels were coded A through D and then arranged by pairs, AB through DC, into 11 different blocks of 12 each. In this sequence, the first letter of each pair designated that passage to be used as a cloze test, and the second letter identified that passage to be used as a measure of information gain (see Table 4). Ten such blocks, one for each grade level (3-12), were made at each of the 11 different difficulty

TABLE 4

BORMUTH'S COUNTERBALANCED DESIGN USED IN STUDY TWO

AB	AC	AD
BC	BD	CD
BA	CA	DA
CB	DB	DC

levels resulting in 1320 test booklets. To ensure average difficulty, the five separate forms of each cloze test were randomly assigned to all of the ten blocks at each of the 11 difficulty levels. In employing the above design, Bormuth thereby assured himself of having all combinations of both test types at all 11 difficulty levels equitably administered to all ten grades.

In analyzing the data collected from this study by means of regression analysis, Bormuth reported the following: 1) Regarding the necessity for calculating separate cloze criterion scores based on information gain for students at each grade level, findings were ambiguous. Although a strong interaction did occur between the cloze test and grade level when the willingness-to-study ratings were used as the dependent variable, there was only partial evidence that came from a small ancillary study that suggested the same was true for information gain. In the

main study, Bormuth (1971) reported that the interaction between cloze scores and information gain scores was ambiguous and may have been made so by the fact "that there seemed to be a very large carry-over effect from administering the same test as both a pre-reading and post-reading test" (p. 77). Therefore, although there was some evidence that older students gained more information at each cloze level, the question still remained unresolved.

2) Regarding the effects of using many passages having multiple difficulty levels, findings from Study Two indicated that although tests with relatively few items had been used, the shapes of the regression were most likely very nearly the same. Again, however, findings were suspect because of the carry-over effect resulting from administering the same test as both a prereading and postreading test. Because of this, Bormuth (1971) indicated that the findings from Study Two only provided "indirect evidence that the regression identity assumption held to a reasonable degree for the information gain regressions" (p. 73). Nevertheless, Bormuth continued, "there seemed to be adequate grounds for proceeding on the assumption that the regression identity assumption was satisfied to a reasonable degree of approximation by the tests employed in this study" (p. 78).

Study Three

In Study Two, carry-over effects caused from administering the same test as both a prereading and postreading

comprehension test resulted in less than clear findings with regard to the shape of the regression obtained over many passages of varying difficulty. Additionally, questions were raised with regard to both the preference ratings that had been gathered as well as the completion-type comprehension tests that had been constructed. As a result, the third study was launched.

Because the basic design used in Study Two had been highly efficient, it was retained in Study Three with three changes: First, instead of using only two passages of matched difficulty, one for the cloze test and the other as a basis for constructing both prereading and postreading completion-type comprehension tests as well as a preference rating scale, four separate passages of equal difficulty were used. From the first passage was constructed the cloze test. From the second came the prereading comprehension test. From the third was taken the postreading comprehension test. And finally, the fourth passage was used as a basis for administering an expanded preference scale in which each student was required to rate a passage on ten separate scales involving subject-matter preference, style, difficulty, willingness-to-study, and use for which the reading material was intended, i.e., as a textbook, as reference material, or as voluntary reading. According to Bormuth (1971), four different passages of matched difficulty were used instead of two, which had been the case in Study Two, "in order to eliminate carry-over

effects that resulted from administering the same test as both a pre-reading and a post-reading test within the short time spans involved in these testing operations" (p. 81).

The second change in Study Three involved using passages believed to be more representative of instructional materials, and the last change saw the implementation of both longer reading selections and comprehension tests in order to improve reliability.

In conducting Study Three, Bormuth selected 1600 students enrolled in grades three through 12 in a middle-class suburban school system. In preparing materials, eight sets of reading passages representing eight different difficulty levels were constructed. Contained within each of the eight different sets were four reading selections that had been matched previously by Bormuth for cloze difficulty. Unlike the 44 passages used in Study Two, these 32 passages were longer (each passage contained a minimum of 250 words) and represented material specifically selected by Bormuth because they were representative of either some randomly selected section of a textbook or some other form of instructional materials. As had been the case in the previous two studies, five different forms of the cloze test were constructed over each of the passages. As had been done in Study Two, a completion-type comprehension test was also made from each passage. Instead of ten short-answer completion items, however, 20 items were constructed. Except for minor

modifications, the manner in which the completion-type test items were constructed was roughly similar to that used in Study Two. Minor changes were made, however, in order to improve the objectivity and the reliability of the test writing procedures. Because Bormuth's work on rating scales in this study is not of significance to this researcher's investigation of cloze criterion scores and information gain, the details of that portion of Study Three will not be treated. Suffice it to say, however, that each student also was required to rate a reading passage on the ten preference scales mentioned above (see p. 60).

Using the above materials, Bormuth constructed his test booklets. Unlike the booklets constructed in Study Two, which utilized only two matched passages per booklet, each test booklet in this study contained four passages, all of matched difficulty. Contained in order within each test booklet were a cloze test made from one of the passages; a completion test made from a second passage (only the test, however, and not the passage was included in the booklet); a third passage followed by the ten preference rating scales; and finally, a fourth passage followed by a 20-question, completion-type comprehension test based on information found in that fourth reading passage.

In constructing the test booklets, the four matched passages within each of the eight difficulty levels were

first coded A through D and then ordered using a Latin squares design. The first passage in each order was administered as a cloze test, the second as a prereading test, the third was rated on the preference scale, and the fourth was administered as a postreading test (see Table 5). Five versions of each block were made, one for each form of the cloze test, which resulted in 20 test booklets at each of the eight difficulty levels. One set of 160 booklets constructed at each of the ten grade levels resulted in the production of 1600 test booklets. According to Bormuth (1971), each student "was completely nested with all factors and with no replications. Thus, there were 160 students for each grade level and 200 students for each difficulty level" (p. 86).

TABLE 5

BORMUTH'S LATIN SQUARES DESIGN USED IN STUDY THREE

A	B	C	D
D	A	B	C
C	D	A	B
B	C	D	A

In analyzing the findings of Study Three, Bormuth generally used the same procedures that had been employed in Studies One and Two, i.e., various adjustments were made to the scores and the data were subjected to regression analysis. With regard to the relationship between cloze scores and information gain, as reflected by a completion-type comprehension test, Bormuth's findings in this study revealed that the three changes in his design (see p. 60) resulted in regression curves that conformed fairly closely to the shape of the regression obtained in Study One (see Figure 1). Had Bormuth's test writing procedures used in Study One been as tightly controlled as those used in Studies Two and Three, and thereby more fully complied with his instructional conformity and behavioral consistency assumptions, the shapes of the regression curves between the two studies would undoubtedly have been even more similar, if not identical. The fact that the regression curves between Studies One and Three were very similar, and could have been made even more similar, is most significant because it confirms that regression curves computed over many passages varying widely with regard to difficulty are nevertheless similar to those computed over only one or two passages of equal difficulty. This, of course, substantiates Bormuth's regression identity assumption, which says that the regression curves must not change regardless of the number of passages used, and thereby indicates that cloze criterion

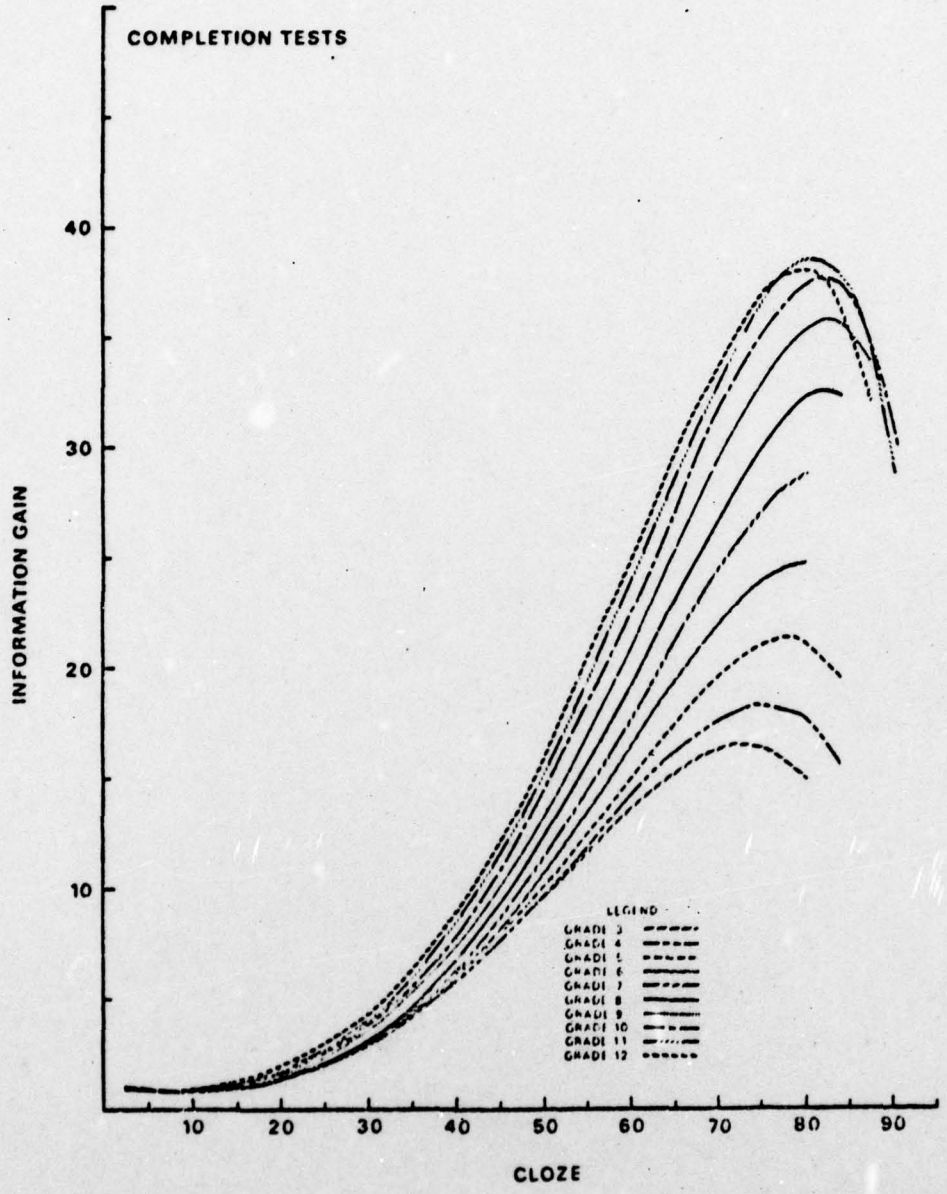


Figure 1. Regression of information gain scores on cloze scores achieved from Bormuth's Study Three.

scores, in fact, can be established. Equally significant is the fact that with the regression identity assumption having been confirmed, subsequent research need not be conducted over many passages having varied difficulty levels. Instead, such research can focus on only one or two passages, as was done in Bormuth's Study One, and be assured that any findings can be generalized to all passages at all different levels of difficulty.

Nevertheless, Study Three only substantiated one half of Bormuth's regression identity assumption, i.e., that regression curves must not change regardless of the passage from which they are made. As Figure 1 shows, the shape and slope of the upper end of the regression curves obtained from Study Three differed for each grade level. For Bormuth, whose objective was to establish maximum comprehension cloze criterion scores gathered from several grade levels of school children, many of whom had not fully developed the reading skill, the findings of Study Three suggested that different criterion scores would have to be established for each grade level. The fact that the second half of Bormuth's regression identity assumption was not substantiated in Study Three is not of great significance for this project, however, which is intended to establish at which point significant comprehension of German prose reading selections begins for adult nonnative readers of German. The reasons for its insignificance are twofold:

First, whereas Bormuth's research dealt with school children, many of whom had not fully developed the skill of reading, this research project is limited to adults, who although reading German at a number of different skill levels ranging from beginning to most advanced, are all nevertheless fully developed and mentally mature readers. This is significant because in analyzing the regression curves of students in grades ten through 12 in Figure 1, it is clearly evident that the shapes of the regressions of these mature young people draw much closer together. This suggests, of course, that the regression identity assumption is, in fact, fully met for adults or near adults who have developed their reading skills.

Second, whereas Bormuth is directing his attention to maximum comprehension, which is reflected in the higher cloze scores, where the variance between grades is at its maximum, this study has been designed to examine the lower scores, where the variance between grades is greatly reduced. Therefore, taking into consideration the fact that this study is using only adult readers and is examining their scores primarily in the lower half of the spectrum, it should be evident that only the first half of Bormuth's regression identity assumption need be considered in this research, and that one cloze criterion score can indeed be established for all adult nonnative readers of German--a cloze score that is valid for all German prose reading selections regardless of the reader's personal reading

skill or the difficulty of the selection.

In examining the findings of Study Three with regard to establishing that point at which comprehension begins to occur, results were slightly different, but probably more accurate, than those obtained from Study One. While Study One suggested that 25 percent on the cloze test might be the point at which comprehension began to occur, Study Three, which was much more comprehensive and operationally replicable, suggested that students had to score at least 35 percent before comprehension began to occur.

Although this author's research project focuses on nonnative adult readers of German and employs a slightly different design incorporating more objective test writing procedures, it is nevertheless expected that a criterion score somewhere near 35 percent will also be established for this research if, in fact, the three research questions stated in Chapter I (see p. 28) can be affirmatively answered.

Summary

The purpose of Chapter II has been to examine that literature related either directly or indirectly to the cloze procedure when it is used as a measuring device for establishing a meaningful and objective cloze criterion score that reflects that point at which reading comprehension begins to occur for nonnative readers of German prose. In carrying out this objective, the examination of literature has progressed through three general

stages:

In stage one, the review began with a very general investigation of the cloze test as it is used as an all-purpose measuring device. Because so much research has been done in recent years involving the cloze procedure as a device for measuring a variety of skills, several comprehensive reviews focusing on the entire spectrum of cloze research were cited (see p. 32).

In stage two of this review of the cloze procedure, the examination was narrowed to those investigations involving the cloze test as a measuring device to be used with foreign language research. During this phase of the review, four interesting things were discovered: First, especially in recent years, most cloze research carried out with foreign languages has been intended to measure general language proficiency rather than specific language skills. Nevertheless, limited work in foreign language readability and reading comprehension work was detected and cited. Second, most foreign language cloze research has been done in English involving speakers of other languages. Third, hardly any cloze research has been conducted involving native English speakers using a foreign language. Fourth, although Anderson (1976) laid some important groundwork for foreign language cloze research and reading comprehension, no research was detected using the cloze procedure to assess any kind of reading skills on nonnative speakers

of German.

In stage three, the final area of cloze research to be reviewed, attention was focused directly on the topic of this research investigation, i.e., using the cloze procedure as a device for establishing objective and meaningful reading criterion scores. In this third area of investigation, various early approaches to the problem were first discussed. Most significant, however, was the research reported by Bormuth (1971) in which a series of five studies was conducted designed ultimately to establish a rational cloze reading criterion score for English-speaking school children. Because of their direct bearing on this author's research project, only the first three of Bormuth's five research projects were dealt with in this review. However, because these three studies form the backbone of this research project, they were dealt with in considerable detail.

CHAPTER III

PROCEDURES

Chapter III is divided into five major sections: 1) a general discussion of the design employed to gather the data; 2) an in-depth discussion and description of the rules developed and applied in order to ensure that both testing instruments would be operationally replicable; 3) an examination of all the details involved; 4) an explanation of the criteria established for scoring the findings; and 5) a brief discussion of the statistical procedures used to analyze the data.

Design

The purpose of this research study is to establish an objective and meaningful cloze criterion score that reflects that cloze level at which college-level, non-native readers of German prose begin to glean substantial information as a result of reading. The value of establishing such an objective cloze criterion score, of course, is that it thereby enables teachers of German to rationally and objectively identify reading selections that are of an appropriate reading difficulty level for their students.

The objective is to be accomplished by comparing and then establishing the relationship of scores carefully matched, college-level, nonnative speakers of German achieve on a cloze test with the scores other matched, college-level, nonnative speakers of German achieve on a specially prepared, operationally replicable, short-answer completion test constructed over the same piece of German prose. Because neither test was, nor could be, manipulated in any way by the researcher, this investigation is of necessity descriptive research and according to Kerlinger (1973), is necessarily ex post facto in nature.

Design Components

The specific design employed in this investigation was inspired by Bormuth's series of research studies conducted in 1971 in English. The simplicity of the design is directly attributable to Bormuth, who,as a result of findings achieved primarily in Study Three (see p. 59), was able to substantiate the regression identity assumption that regression curves computed over many passages of varying difficulty were nevertheless similar to those computed over only one or two passages of equal difficulty. As a result, the design used in this study could be simplified to include only two reading passages, one to match the students according to reading ability, and one to gather the actual data for analysis, and nevertheless still be assured that findings could be generalized to all passages at any difficulty level, a requirement

necessary for establishing meaningful and generalizable cloze criterion scores.

In collecting the data used in this research project, the following steps were accomplished: First, because researchers (Bormuth, 1962 and Fletcher, 1955) had demonstrated the cloze procedure to be both a valid and reliable measure of reading comprehension ability, the students selected to participate in the study were initially given a cloze test constructed from a reading selection of German prose in order to form pairs of readers matched according to reading ability.

Second, a second selection of German prose was prepared in two different versions for testing: 1) Three different forms of the cloze test were constructed in which different combinations of every eighth word were deleted. For example, on one cloze form, words 2, 10, 18, etc.were deleted; on a second form, words 4, 12, 20, etc. were deleted; and on the final cloze form, words 5, 13, 21, etc. were deleted. Multiple forms of the cloze test were used to control for sampling error and ensure that the actual difficulty of the passage was being sampled and tested. 2) A specially prepared, operationally replicable, short-answer completion test was constructed using guidelines outline in this chapter (see p. 76).

Third, one member from each matched pair was randomly selected and administered the specially prepared completion test, while the other matched-pair member was administered

one of the three forms (randomly chosen) of the cloze test (see Figure 2). The scores achieved from these two test types were then compared in order to determine what kind of relationship existed, if any.

Instrument Development

In order to establish meaningful and objective cloze reading criterion scores based on information gain, it is necessary to compare the relationship of cloze test scores and objective test scores. For this comparison to be valid and thereby generalizable to all German prose reading selections, however, it is essential that both Bormuth's instructional conformity and behavioral consistency assumptions be met. To meet these two assumptions, both the cloze test and the objective test must be prepared in an operationally replicable manner. By their very nature, cloze tests are replicable because they are prepared by systematically deleting words according to a set of predetermined rules. Objective tests, however, are normally generated in a very subjective manner, being greatly influenced by the individual biases of each test writer. An integral and essential part of this research project, therefore, was to develop an objective test writing algorithm in which a set of rules was generated for constructing an operationally replicable, short-answer, completion-type objective test in German. In this chapter rules used for generating both the objective test and the cloze test will be set forth.

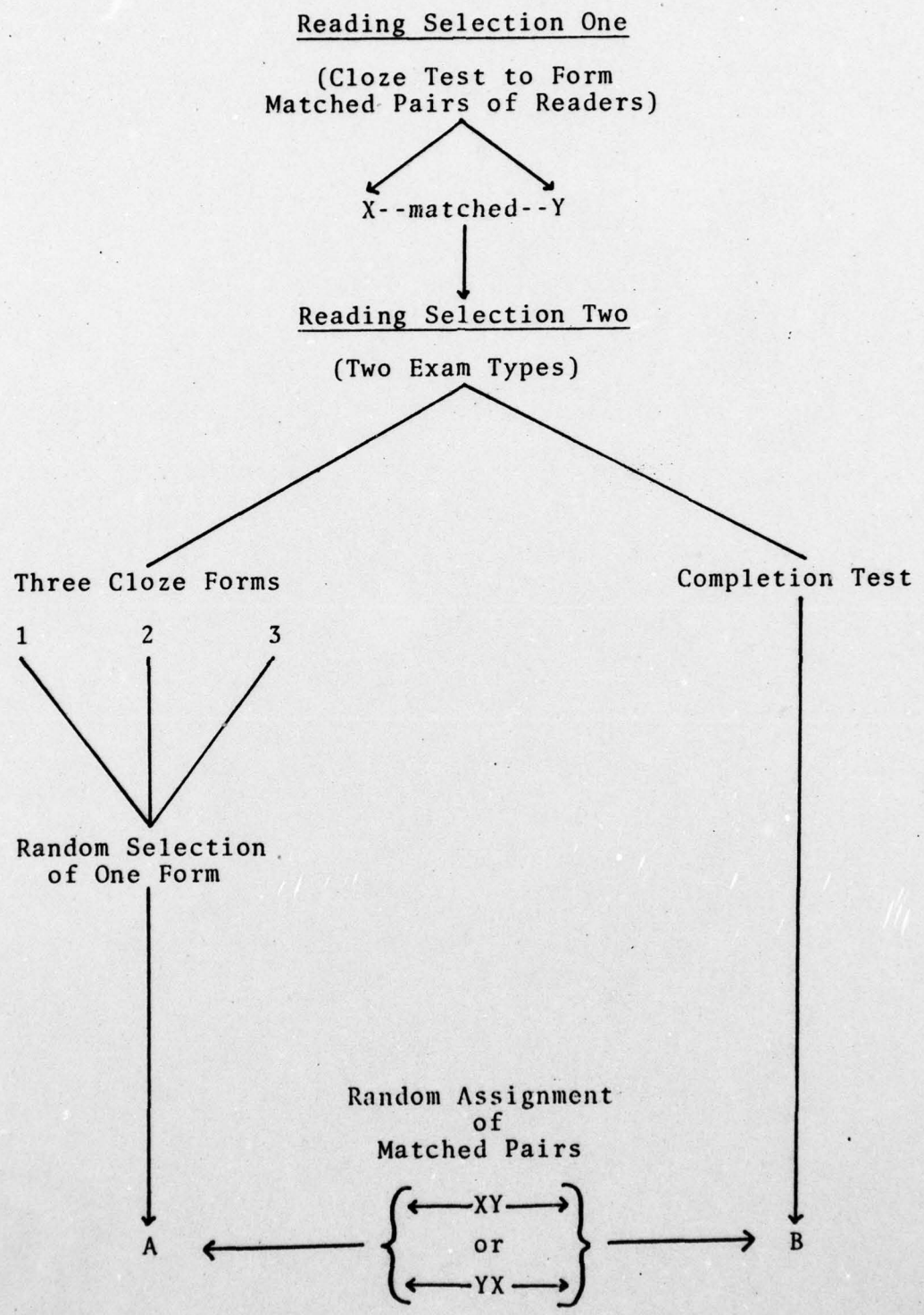


FIGURE 2. Research design used to gather data for analysis.

The Objective Test

In developing the following algorithm for writing replicable, completion-type objective tests, initial inspiration was drawn from Bormuth (1970), who developed test writing procedures for English texts based on generative transformational grammar. The actual test-writing algorithm used in this study, however, was patterned after ideas suggested by Finn (1973), who was inspired by research conducted by Stockwell, Schachter, and Partee (1968), in which the theories of both generative transformational grammar and case grammar were integrated to develop a very detailed grammar of English.

The algorithm used in this study contains 33 test-writing rules divided into four general sections: 1) rules for sentence selection, 2) rules for sentence preparation, 3) rules for sentence analysis, and 4) rules for question preparation. In presenting the rules of this test-writing algorithm, each rule will be stated and then followed by an explanation and/or example, where necessary.

Rules for Sentence Selection

A key first step in generating an operationally replicable test-writing algorithm is to determine which sentences within the chosen reading selection contain information or ideas of such significance to the passage that their content should be considered for inclusion as one of the test questions. To make such a determination, the

following rules must be applied:

Rule 1. Determine each word's Text Frequency (TF) in the reading selection by counting how often it occurs in the passage.

Explanation and Example. If there were an interest in writing a completion test over a 500-word German selection about science, Rule 1 dictates that one simply would first count how often every word in the passage occurred. The frequently used definite article "der," for example, might occur 25 times in the selection and therefore have a Text Frequency of 25, while the noun "Atom" might occur only five times and therefore have a Text Frequency of only five.

Rule 2. Compute an Interim Value (IV) for each word in the passage by dividing each word's Text Frequency by the total number of words (TW) contained within the reading selection.

Explanation. With "der" occurring 25 times in the 500-word reading selection about science, it would be given an Interim Value of .05, while "Atom," occurring only five times, would be assigned an Interim Value of .01.

Example.

$$\text{IV for "der"} = \frac{TF}{TW} = \frac{25}{500} = .05$$

$$\text{IV for "Atom"} = \frac{TF}{TW} = \frac{5}{500} = .01$$

Rule 3. Determine the Meier Frequency (MF) of all words, i.e., how frequently each word occurs in a count of nearly 11 million (10,910,777) running German words, by consulting Meier's Deutsche Sprachstatistik (1967).

Explanation. Utilizing the definitive work Häufigkeits-wörterbuch der deutschen Sprache compiled at the end of the 19th Century by F. W. Kaeding, Meier has alphabetically listed every form of every German word that appears ten or more times in Kaeding's work. Meier's work consists of 41,083 separate word forms and lists how frequently each occurs in Kaeding's count. Meier's list constitutes 96.3 percent of all words found in running German texts.

Example. "Der" has a Meier Frequency of 341,522. "Atom," on the other had, has a Meier Frequency of only 14.

Rule 4. In order to determine each individual word's importance to passage understanding, compute the Significance Coefficient (SC) of all words in the passage by dividing each word's Interim Value by its Meier Frequency. If a word is not listed by Meier, automatically assign that word a Meier Frequency value of nine (9).

Explanation. The intent of Rule 4 is to identify those words truly important to passage understanding and to prevent words such as articles and prepositions, which always occur frequently, from being selected as significant to passage understanding just because they occur frequently.

Example.

$$\text{SC for "der"} = \frac{IV}{MF} = \frac{.05}{341,522} = .14 \quad (.00000014)$$

$$\text{SC for "Atom"} = \frac{IV}{MF} = \frac{.01}{14} = 714.28 \quad (.00071428)$$

Rule 5. Select and list alphabetically by Significance Coefficient the 50 most significant words in the passage.

Explanation. That word with the highest Significance Coefficient will be listed as word one; the word with the next highest Significance Coefficient will be listed as word two, etc., until the 50 most significant words have been identified. In those cases where there are two or more words with the same Significance Coefficient, they will be arranged alphabetically and numbered consecutively.

Example. See Appendix C, p. 173.

Rule 6. Locate and list numerically those sentences in the reading selection in which each of the 50 most significant words occurs. These are the sentences that will be used in generating questions.

Explanation and Example. If "Atom" had the highest Significance Coefficient in a 500-word passage on science and occurred five times in the passage, question one would be derived from the first sentence in the selection in which "Atom" occurred. Question two would be derived from the second sentence in the passage in which "Atom" occurred, etc., until all five sentences containing the word "Atom" had been utilized in generating questions.

Rules for Sentence Preparation

After identifying those sentences within a reading selection that contain words key to passage understanding, steps must be taken to modify these sentences into their basic semantic components in preparation for structurally analyzing them. To modify selected sentences into their basic semantic components, the following rules must be applied:

Rule 7. Replace all pronouns with the appropriate noun.

Example. "Dort lebte er auf einer Oase ganz wie in England." becomes "Dort lebte der Engländer auf einer Oase ganz wie in England."

Rule 8. Replace adverbs such as "dann," "später," "dort," etc. with the appropriate phrase or clause.

Example. "Dort lebte der Engländer auf einer Oase ganz wie in England." becomes "In der Wüste lebte der Engländer auf einer Oase ganz wie in England."

Rule 9. Treat all compound sentences joined by "und" as separate sentences ignoring the conjunction.

Example. "Der Engländer badete fleissig und trank viel Tee." becomes "Der Engländer badete fleissig." "Der Engländer trankviel Tee."

Rule 10. If an adverb/adverbial (ADV) is present in a sentence analyzed as more than one sentence, repeat the ADV in all sentences.

Example. "Jeden Tag badete der Engländer fleissig und trank viel Tee." becomes "Jeden Tag badete der Engländer fleissig." and "Jeden Tag trank der Engländer viel Tee."

Rule 11. Rewrite sentences with elliptical constructions restoring the word or words that complete the construction.

Example. "Guten Tag." becomes "Ich wünsche Ihnen einen guten Tag."

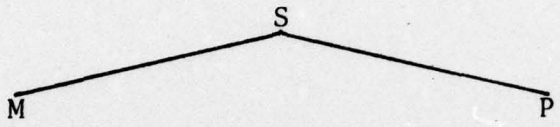
Rules for Sentence Analysis

Having selected and modified into their basic semantic components those sentences that contain words and information key to passage understanding, sentences are now

ready for structural analysis. In structurally analyzing sentences, the following rules must be followed:[1]

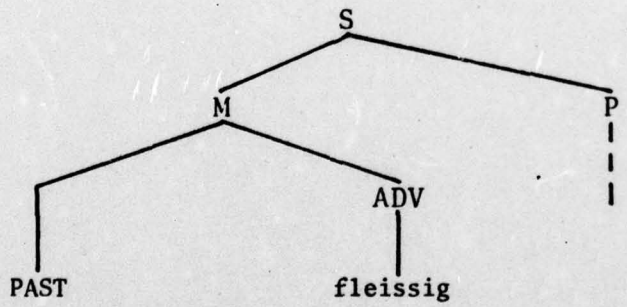
Rule 12. Divide each prepared sentence (S) into modality (M) and proposition (P).

Example.



Rule 13. Under modality, place negative elements (nicht); adverbs (sofort, ganz, schnell, etc.); dependent clauses functioning as sentence adverbials (weil der Mann sehr krank ist,....etc.); and other elements understood as modalities on the sentences as a whole (tense; voice, other than active; use of modal verbs).

Example. Following Rule 13, in the sentence "Jeden Tag badete der Engländer fleissig.", modality would be constructed in the following manner:

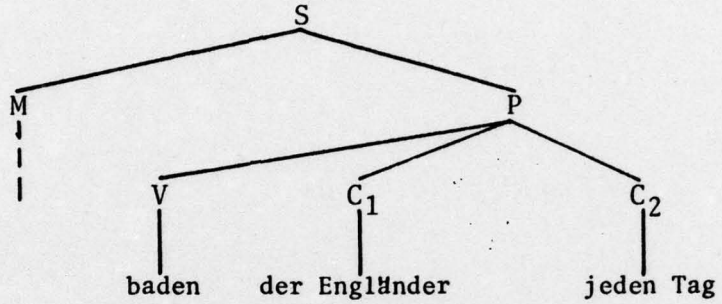


Rule 14. Under proposition, place the verb (V) plus phrases having case relationships (C) to the verb.

---

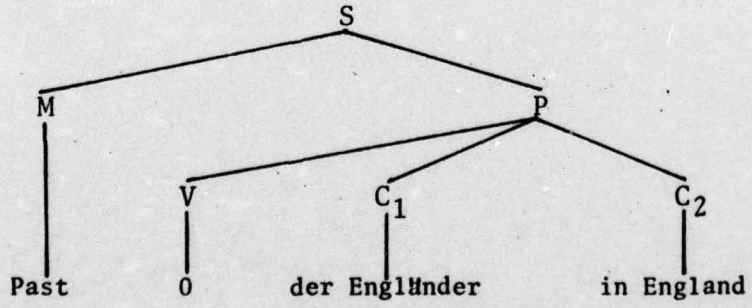
[1]Examples provided with the following rules are, of necessity, very elementary. For more detailed constructions and analyses see Appendix D, p. 178.

Example. Following Rule 14, in the sentence "Jeden Tag badete der Engländer fleissig.", the following elements would be placed under proposition:



Rule 15. Do not treat "sein" as a verb. If one of its conjugated forms is followed by a predicate adjective, enter the predicate adjective as the verb. Otherwise, assign any form of "sein" the value of zero (0).

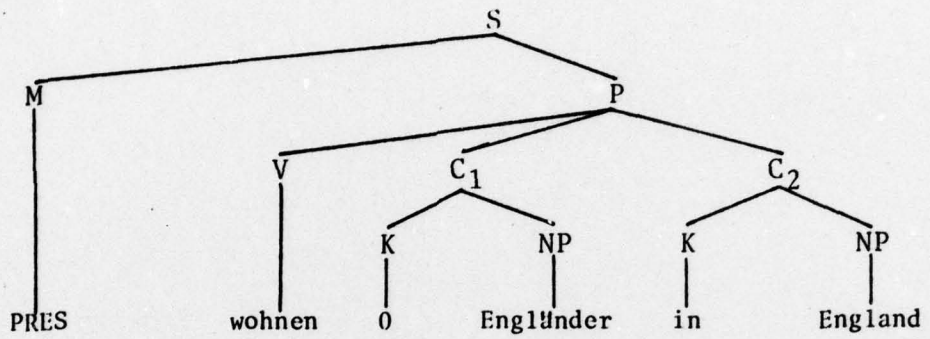
Example. Following Rule 15, in the sentence "Der Engländer war in England.", the verb form "war" would be given the value of zero:



Rule 16. Analyze all case relationships (C) as Prepositions (K), which indicate the relationship of the phrase to the verb, and noun phrases (NP). K may be, and often is, zero (0).

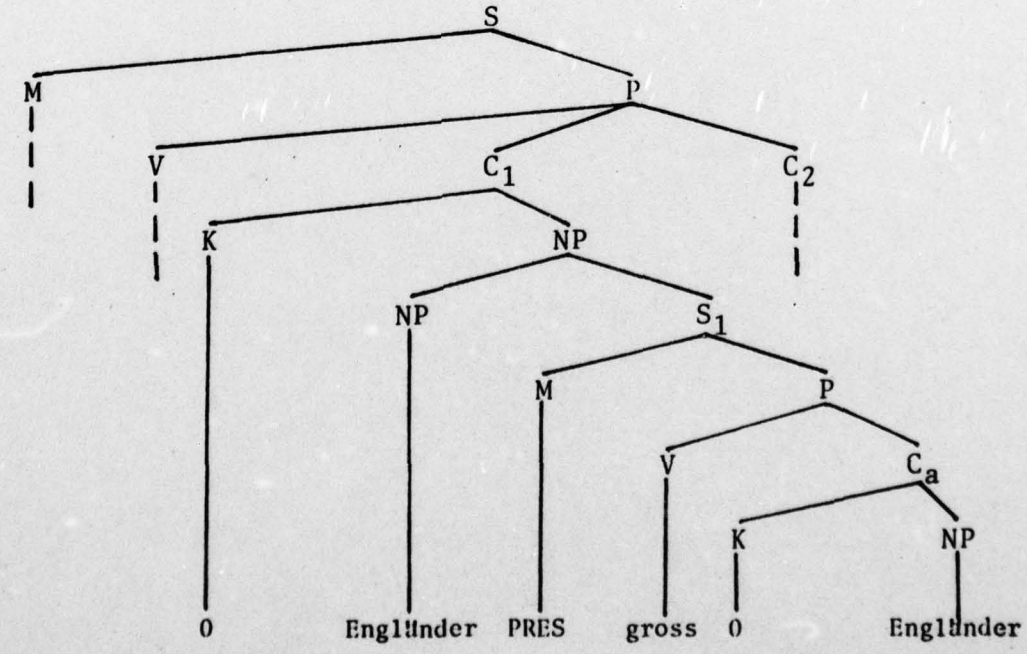
Example. Following Rule 16, in the sentence "Der Engländer wohnt in England.", the case relationships (C)

under the proposition (P) would be further broken down as follows:



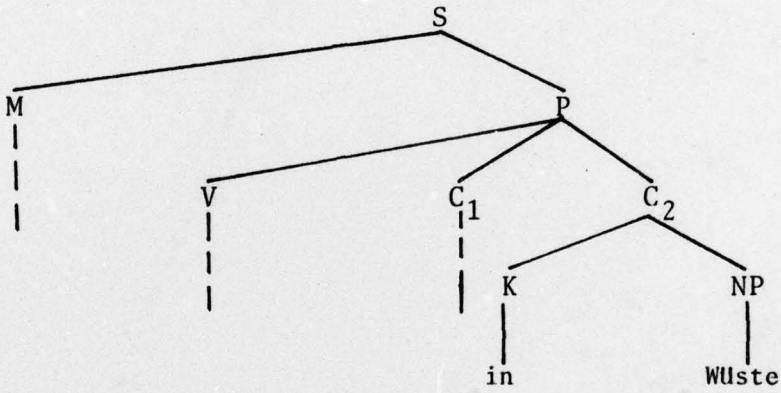
Rule 17. Construe all relative clauses (regular and reduced) as sentence (S)--right sister to the noun phrase (NP) and under the noun Phrase (NP).

Example. Following Rule 17, in the sentence "Der grosse Engländer wohnt in England.", the reduced relative clause "Der grosse Engländer would be treated as follows:



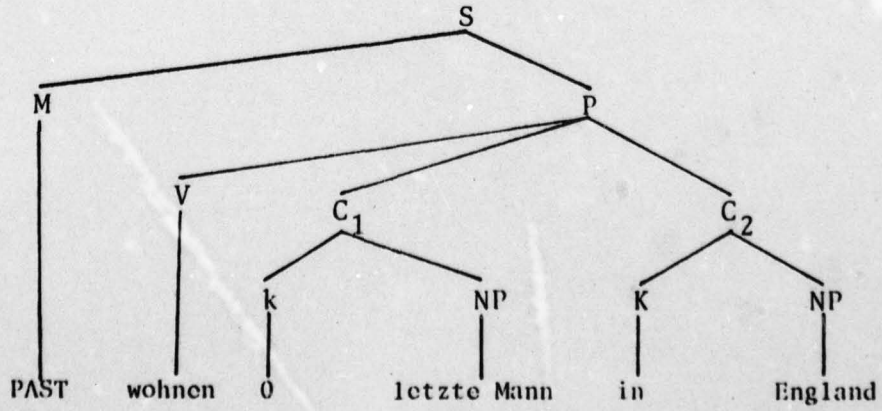
Rule 18. When a phrase appears to be analyzable either as a reduced relative clause under NP or as an adverbial under C, as in "Der Engländer studierte das Kamel in der Wüste.", analyze the phrase as C.

Example. Following Rule 18, in the sentence "Der Engländer studierte das Kamel in der Wüste.", the phrase "in der Wüste" would be treated as follows:



Rule 19. Do not analyze ordinals (der erste, der nächste, der letzte, etc.) or quantifiers (einige, viele mehrere, etc.) as S under NP. Instead, the head noun and these modifiers are to be treated as an inseparable unit under C.

Example. Following Rule 19, the sentence "Der letzte Mann wohnte in England." would be analyzed as follows:

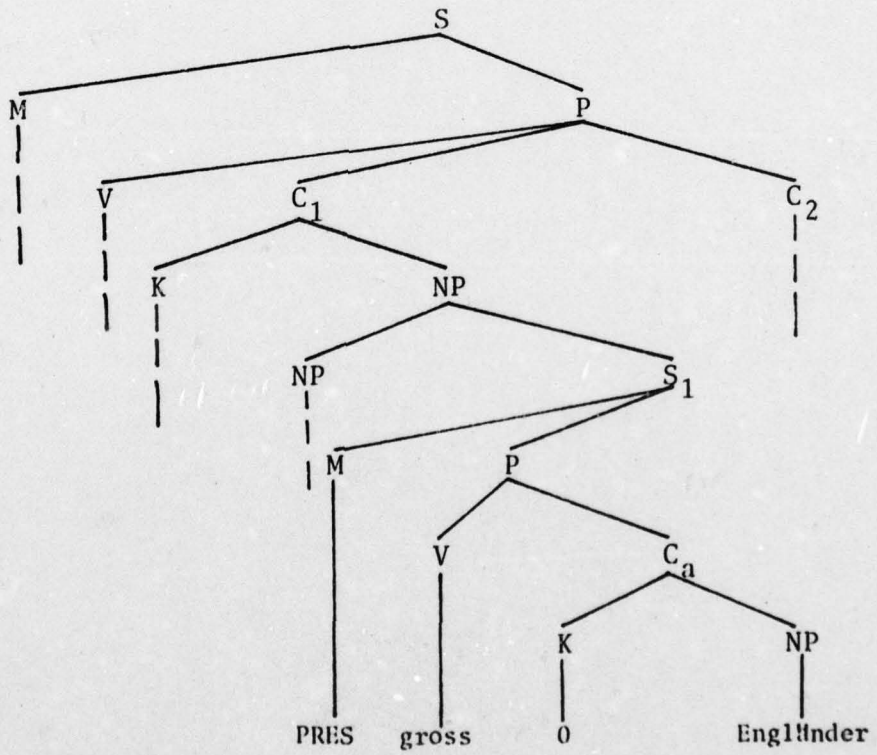


Rule 20. Treat all problem modifiers not dealt with in Rules 17, 18, and 19 as part of the head word.

Rule 21. Allow only S to be the daughter of NP.

Explanation. Rule 21 compliments and supports Rule 17 by allowing only for an S to be directly attached to an NP.

Example. Following Rule 21, in the sentence "Der grosse Engländer wohnt in England.", the reduced relative clause "der grosse Engländer" must be represented as an S under NP as follows:

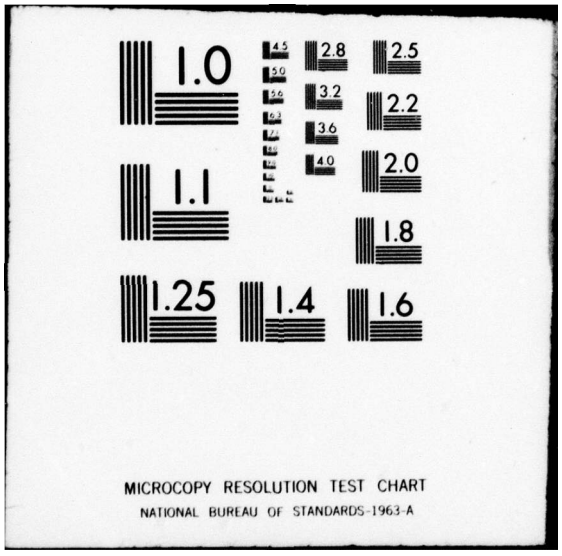


Rule 22. Analyze conjoined constituents as compound constituents under C.

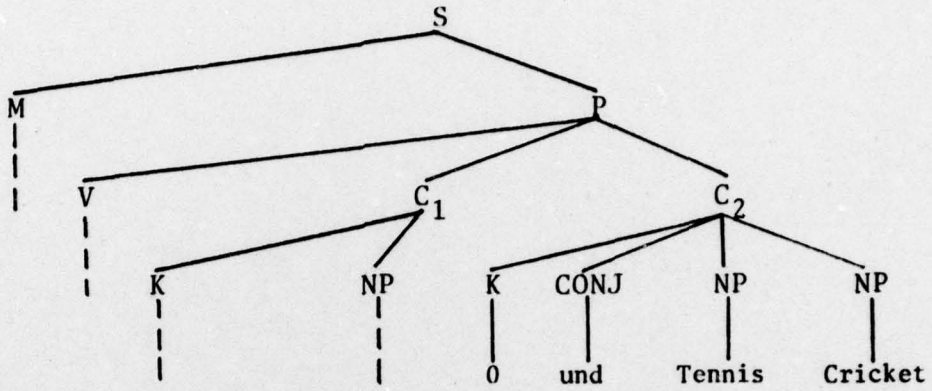
Example. Following Rule 22, in the sentence "Der Engländer spielte Tennis und Cricket.", the conjoined

AD-A065 489 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO F/G 5/10
THE CLOZE TEST AS A PROCEDURE FOR ESTABLISHING OBJECTIVE GERMAN--ETC(U)
1978 B M STRONG
UNCLASSIFIED AFIT-CI-79-100D NL

2 OF 3
AD
A065 489

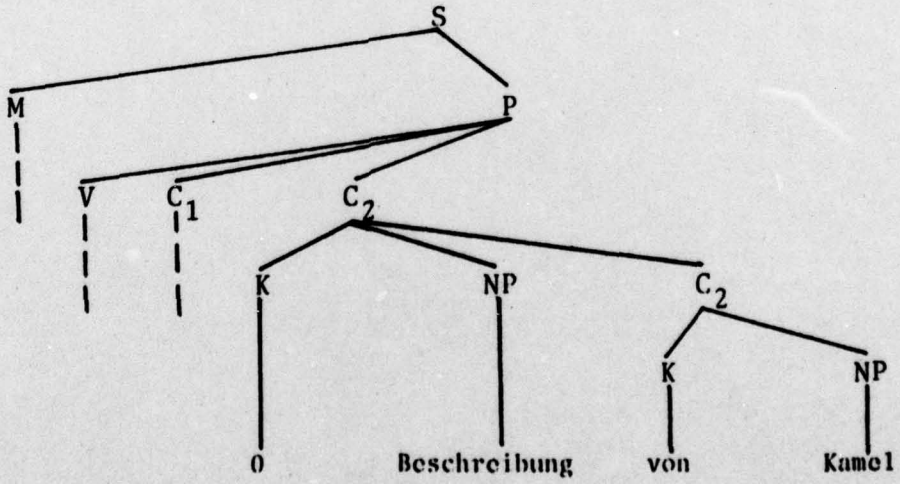

1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

constituents "Tennis und Cricket" must be represented as follows:



Rule 23. Analyze genitive constituents as extensions of the C to which they refer.

Example. Following Rule 23, in the sentence "Der Engländer stellte eine Beschreibung des Kamels zusammen.", the genitive "des Kamels" must be represented as follows:



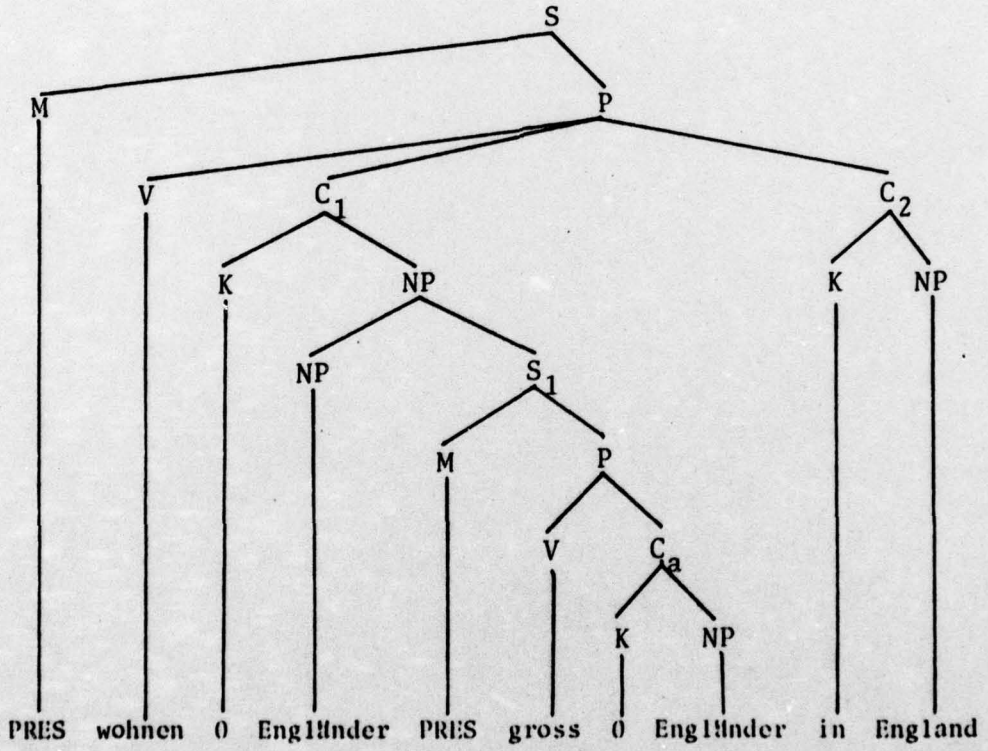
Rules for Question Preparation

Having selected, modified, and structurally analyzed those sentences containing words and information key to

passage understanding, steps now may be taken to generate actual test questions. In transposing sentences into questions, the following rules must be applied:

Rule 24. In selecting that sentence constituent to be replaced with an interrogative form, consider only case constituents labeled $C_1 \ldots C_n$ and sentence constituents labeled $S_1 \ldots S_n$.

Example. Following Rule 24, in the sentence "Der grosse Engländer wohnt in England.", only constituents C_1, C_2, and S_1 may be considered for replacement with an interrogative form:



Rule 25. Determine the constituent to be deleted by means of random selection.

Rule 26. If either a C_n or S_n constituent, which has subordinate S_n constituents, is selected for replacement,

delete the subordinate S_n constituents as well.

Explanation. Using the example presented under Rule 24, Rule 26 directs that should constituent C_1 be randomly selected for replacement with an interrogative form, constituent S_1 would be deleted as well because it is subordinate to C_1. If S_1 should be randomly chosen, however, constituent C_1 would remain because it is superordinate to S_1

> Rule 27. In each analyzed sentence, replace the randomly deleted constituent with one of the following interrogative forms: "wann," "warum," "was," "was für," "was + tun," "welcher," "wer," "wie," "wie lange," "wo," "woher," "wohin," "wo(r) + Preposition."

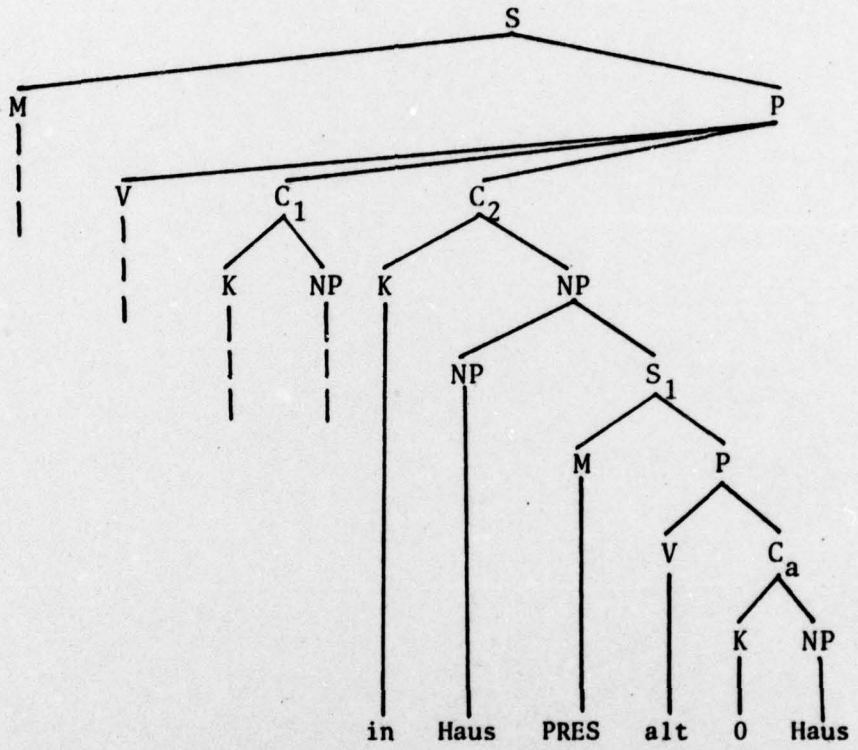
Example. Using the sentence presented under Rule 24, had C_2 (in England) been randomly selected for interrogative replacement, semantic meaning would require that the interrogative "wo" be selected from among the choices listed in Rule 27 to generate the question.

> Rule 28. If in generating a question a C_n is randomly selected for interrogative replacement, place the appropriate interrogative to the front of the sentence and follow German inverted word-order rules.

Example. Using the sentence presented under Rule 24, had C_2 (in England) been randomly selected for interrogative replacement, the question would read as follows: "Wo wohnt der grosse Engländer?"

> Rule 29. If in generating a question an S_n is randomly selected for interrogative replacement, follow Rule 28 unless the S_n to be deleted is subordinate to a C_n containing a K with a value other than zero (0). In such instances, begin the question with the K followed by the semantically appropriate interrogative.

Example. Following Rule 29, if S_1 were randomly selected for interrogative replacement in the sentence "Der Engländer wohnt in einem alten Haus.", represented below, the resultant question would read as follows: "In was für einem Haus wohnt der Engländer?"



Rule 30. Although the indefinite pronoun "man" and its declined forms "einen" and "einem" may serve in a C_n capacity, do not consider them for deletion.

Rule 31. Should selection procedures involving two adjacent questions result in the same answer for both questions, randomly select another constituent for deletion from the latter question.

Rule 32. Should any constituent chosen for deletion prove to be stylistically unacceptable, eliminate that constituent from consideration and randomly select from the remaining constituents.

Rule 33. After a question has been generated from a sentence, do not reuse the sentence a second time even though it may contain one or more additional significant words.

The Cloze Test

Unlike the objective test, which is most difficult to make operationally replicable, the cloze test generally presents no difficulties. In fact, according to Rankin (1965), "one of the most characteristic features of the cloze procedure is the use of a completely mechanical system for choosing words to be deleted" (p. 144). Being "completely mechanical," of course, implies operational replicability. Therefore, in constructing a cloze test, the primary task is not in ensuring the cloze test will be operationally replicable, but rather in determining what form of replicability the test will take, i.e., what rules will be followed in mechanically deleting words.

Two Types of Cloze

In constructing cloze tests, the two most commonly used systems for choosing words to be deleted are random deletion and an every nth deletion. Because the every nth deletion is easier to construct, most researchers have preferred this method. In selecting an every nth deletion, however, the question is immediately raised concerning what deletion rate to use. For example, in addition to experimenting with random deletion, Taylor (1953) also explored deleting every fifth, seventh, and tenth word. According to Anderson (1976), Anderson (1965) deleted

every eighth word, Weaver (1961) every tenth word, and Greene (1965) every twelfth word. The best controlled research was that of MacGinitie (1961) who compared the effect of deleting every third, sixth, twelfth, and twenty-fourth word and concluded that "additional uninterrupted context beyond five words did not help in the restoration of the missing words" (p. 127). Based on MacGinitie's work, then, suggesting that five-word deletions resulted in statistical independence, most research involving the cloze procedure, including the work of Bormuth, has employed the every fifth-word deletion pattern. As Rankin (1965) points out, however:

> It does not follow, however, that this pattern (every fifth-word deletions) is suitable for all materials, age groups, and purposes for which the cloze procedure may be used. More work is needed on this question (p. 145).

As has already been pointed out in Chapter II, almost all cloze research has been conducted in English using native speakers of that language. However, according to Oller (1973), "there is little if any reason to assume that conclusions from research with native speakers can validly be generalized to the case of non-native speakers" (p. 107). Therefore, although it has been determined with native speakers that deleting more than one out of every five words in a cloze test creates a test so difficult that discriminatory power is lost, and to delete less frequently than one in 12 does not substantially change the quality of the test, such findings do not suggest how

many words should be between deletions when testing non-native speakers of German. All that is suggested is that both for native and nonnative speakers, five-word deletions will be more difficult than ten-word deletions. To be any more specific would result in subjectivity because according to Oller (1973), "the parameter of distance between blanks has not been investigated systematically with non-native speakers" (p. 107).

Justification for an Eight-Word Deletion

Based on a lack of specific research providing deletion guidance for nonnative speakers of German, an eight-word cloze deletion system was selected in this research for two reasons: 1) Because students who possessed varying degrees of German language reading proficiency, ranging from very elementary to near native, were needed in order to determine at which point on the cloze test significant comprehension began to occur, it was felt that the difficulty index of five-word deletions ultimately might be so high as to invalidate inordinately large proportions of the sample and therefore adversely affect the findings. Using eight-word deletions, on the other hand, would probably significantly reduce the difficulty index and likely result in more usable data spread over a broader range of readers. 2) A pilot study involving near-native readers revealed that a five-word deletion system applied to the passages selected for this research was very time consuming for the students and required approximately

30 minutes to complete. Because all data had to be collected in a 50-minute class period, and all readers at all levels needed ample time to complete the selections, it was decided that eight-word deletions would allow less-skilled readers additional time to complete the cloze test. Therefore, the cloze tests used in this research followed every eighth-word deletion patterns.

Study Details

In discussing the details of this project, three areas will be examined: the subjects utilized, the passages employed, and the testing sequence followed.

The Subjects

The sample used in this research was drawn from a target population of college-aged, nonnative students of German who had completed a minimum of two college quarters of formal German instruction at one of two central Ohio institutions (The Ohio State University and Otterbein College). Because the goal was to establish a cloze criterion score by determining at which percentage point on the cloze test nonnative readers of German began to glean substantial information, it was essential that the sample be relatively large and contain proportionate numbers of students whose German reading ability ranged from very elementary through to near native. Therefore, 315 students enrolled in 27 classes differing in difficulty from third quarter first year through to the graduate level were selected. Due to absences, students enrolled in two or

more classes, and students desiring not to participate in the project, 210 students (108 males and 102 females) were ultimately utilized. A completely random selection of the sample from the population was not accomplished for two reasons: 1) only a limited number of nonnative students was available, and 2) a relatively large sample of at least 200 students was desired. Nevertheless, random assignment of students to treatment was accomplished. Because only partial randomization was accomplished, the findings of this study can only be generalized to the larger population of all college-aged, nonnative students of German by that degree to which the sample used in this study is representative of that population. There is no apparent reason to believe, however, that the students involved in this study were significantly unlike those college-aged, nonnative students enrolled at other institutions throughout the country, which suggests that the findings obtained in this project are meaningful and generally can be applied to the larger population.

The Passages

A great deal of forethought was carried out in selecting the two passages used in this research. At least three factors needed careful consideration: First, in order to ensure that the difficulty level of the reading selections would be accurately assessed, it was necessary to select passages of such a length that cloze tests containing at least 50 blanks could be constructed. This determination

was made based on two inputs: 1) a comment by Taylor (1956), who indicates that "a series of about 50 blanks is roughly sufficient to allow the chances of mechanically selecting easy or hard words to cancel out and yield a stable score of the difficulty of a passage" (p. 48); and 2) research conducted by Bormuth (1964), who investigated the variability of passage difficulty at different test lengths, found that differences in difficulty diminished as test length increased, and recommended that cloze tests contain at least 50 blanks.

Second, in addition to having reading selections from which cloze tests of at least 50 blanks could be constructed, finding passages of an appropriate difficulty level was critical. Because it was essential for the sample to contain students possessing reading skills ranging from very elementary to near native in order to determine at which cloze point readers began to gain substantial information, it was also necessary to choose reading selections that were neither too easy nor too difficult. For example, were the passages to prove to be either too easy or too hard, all readers in the sample would achieve either very high or very low scores. However, without a wide range of scores, it would be most difficult to determine a cloze criterion score reflecting that cloze point where information began to be gleaned.

Third, it was decided to use passages based on fiction rather than fact in order to eliminate the need for administering the objective test twice during two separate testing sessions, once as a prereading test and once as a postreading test to determine how much prior knowledge each student taking the objective test had of the material in the reading passage. Using a passage of unfamiliar fiction guaranteed that no one would possess any prior knowledge of the information in the reading selections and thereby would also eliminate the need for a pretest administered during an extra testing session.

Taking the above three factors into consideration (length of passage, difficulty level, and subject matter), two reading passages were selected, one from which a cloze test was made to match students according to reading ability, and one from which both a cloze test (three different forms) and an operationally replicable, short-answer, completion-type objective test were constructed for use in the data analysis. Because of the extreme importance of difficulty level, prior to final passage selection all potential passages were reviewed by an experienced, widely published native-German writer of graded German reading selections. Based on author evaluation of the difficulty level of the work and subsequent agreement and approval by the experienced German writer, "Die Riesen und die Zwerge," a 493-word fairy tale was selected as a basis for constructing a cloze test to match

students according to reading ability, and "Das Kamel," a 572-word fictitious story was chosen from which to generate the cloze and objective tests for use in the actual data analysis. Both stories came from a German graded reader entitled Glück Auf by Müller and Wenckebach (1901). From "Die Riesen und die Zwerge" was constructed a 58-blank cloze test. From "Das Kamel" were constructed three different forms of a 69-blank cloze test and a 38-question, completion-type objective test that was operationally generated (see Appendix B).

Testing

In order to gather the necessary data for analysis, two testing sessions with all participants were required, one for matching students according to reading ability, and one for collecting the data to be analyzed. In preparation, permission was received from 18 university teachers (three teaching assistants, three instructors, and 12 professors) to visit their 27 respective German classes during two separate class periods.

In the first testing session, approximately 40 minutes of class time were required. In order to motivate all participants, the importance of reading appropriately difficult reading selections was first emphasized. Brief written instructions for taking the cloze test were then reviewed, and the test was administered (see Test I in Appendix B). In matching pairs of students according to reading ability, all scores were first converted to a

percentage, ranked according to size, and then matched in order, a pair at a time. In only seven of 105 matched pairs did the cloze scores differ by more than .9 percentage points. In no instance was the difference greater than 3.5 percentage points. As a general indication that a broad range of reading skills was contained within the sample, and that the passage "Die Riesen und die Zwerge" had been of an appropriate difficulty level, the high score achieved was 97.4 percent, the low score was 12.1 percent, and the mean for all 27 classes was 54.9 percent. To visit all classes, score the results, and arrange students into matched pairs according to reading ability required approximately ten days to complete.

In preparation for the second and final testing session, the two members of each matched pair were randomly designated pair member X or pair member Y. For member X, one of the three forms of the cloze test was prepared, the three forms being assigned in such a manner that they were equally distributed throughout the sample. For member Y, a copy of the completion-type objective test was prepared.

During the testing session, students in each class were told initially that they would be taking one of two types of tests that were being administered, either a test similar to the one they had taken in testing session one (a cloze test), or a short-answer completion test to be taken after carefully reading a German short story. The

tests were then disseminated and students were admonished to read all instructions carefully (see Tests IIA and IIB in Appendix B). To administer the tests required approximately 40 minutes of class time.

Scoring Criteria

The Importance of Scoring

A key factor effecting the establishment of an objective and generalizable cloze criterion score based on information gain is the manner in which the data are scored. By varying the scoring procedures, of course, the cloze criterion score is also varied. Therefore, it is imperative to select the most reliable scoring method possible and then ensure that it is consistently administered. In establishing scoring criteria, several factors merit consideration: On the one hand, because German teachers using the cloze procedure to determine the appropriateness of reading selections will have to apply the same scoring procedures used in this research in order to achieve valid findings, it is extremely desirable to employ a very simple set of rules. On the other hand, however, if reliability is sacrificed for the sake of scoring ease, the primary purpose for having established an objective cloze criterion score will have been eliminated. How, then, does one establish the most appropriate method of scoring?

Some Background Information

In reviewing the literature, it is immediately apparent that most cloze researchers of English have

utilized the exact scoring method, a procedure that gives credit only when the reader fills the blank with precisely that word that was originally deleted. Such decisions have been based on the findings of investigators such as Taylor (1953), Rankin (1957 and 1959), Ruddell (1963), and Bormuth (1965), all of whom compared the cloze scores achieved by using the exact scoring method with those established by counting both synonyms and exact matches and, according to Rankin (1965), found that "for measurement purposes, the more tedious and subjective procedure of giving credit for synonyms is not worthwhile" (p. 147). Despite such findings however, subsequent researchers such as Darnell (1968), Bowen (1969), and Oller and Inal (1971), who have experimented with nonnatives taking English cloze tests, have preferred to accept scoring procedures in which at least partial credit was given for contextually acceptable responses. To lend support to such actions when administering the cloze test to nonnative speakers of English, Oller (1972b) gave three cloze tests of varying difficulty to 398 foreign students that had also taken the UCLA ESL Placement Examination (2A Revised). After evaluating each of the cloze tests with five different scoring methods and then correlating the findings with the ESL Placement Examination, Oller (1973) stated that "the data supported the conclusion that with nonnative speakers the method of allowing any contextually acceptable response is significantly superior to the exact

word scoring technique" (p. 109).

With nonnative speakers, then, there exists doubt as to the superiority of the exact method for scoring English cloze tests. Further complicating the issue in this research, however, is the fact that German, not English, is the language under investigation. Unlike English, which is only lightly inflected, German is a highly inflected language. Therefore, in scoring a cloze test, how is one to score a blank that is filled with the correct but improperly inflected word, a problem that seldom occurs for either native or nonnative speakers of English? In order to resolve such problems and determine the most reliable scoring criteria for this research, it was necessary to conduct a pilot study.

The Pilot Study

The specific purpose for conducting a pilot study was to determine the answers to two general questions: 1) Which scoring method is more reliable, the exact or accepting any response that is contextually acceptable and demonstrates comprehension? 2) In dealing with a highly inflected language such as German, should the ability to inflect correctly also be a scoring consideration? Before scoring all the 58-blank cloze matching tests that had been administered in testing session one, the results achieved by two of the 27 classes first were experimented with in order to answer the above two questions and thereby establish the most appropriate scoring

criteria.

In answering question one and determining whether the exact scoring method was better than the acceptable-word method, cloze tests taken by a class consisting of the most advanced German students were chosen for evaluation. The selection of this class, which contained 16 students, was made because it was anticipated that advanced students with large vocabularies would be inclined to fill the cloze blanks with a variety of acceptable words and at the same time make only minimal inflectional errors. Therefore, the cloze results achieved by this class were scored twice using two separate scoring procedures. In order to focus on the effects of the exact scoring method as compared with the acceptable-word method, inflectional considerations were treated as a constant, and during both scorings only half credit was awarded when inflectional mistakes were made. During the first scoring, only the exact word that had been deleted was accepted as correct. If the exact word was selected but improperly inflected, however, only one-half credit was awarded. During the second scoring, any contextually correct word was accepted. Again, however, if an incorrect inflection was used, only one-half credit was awarded. The reliabilities of these two scoring methods were then determined by first correlating odd- and even-numbered test items and then correcting the resultant correlations by applying the Spearman-Brown prophecy formula. By using the exact

scoring method, a reliability of .66 was achieved. Accepting any contextually correct word boosted reliability to .88, however, and suggested that this was a better scoring procedure for this research project.

In answering question two and determining whether at least some inflectional considerations should be included as part of the scoring criteria, a class of 15 first-year German students was selected. In this instance a beginning class was chosen because it was anticipated that unlike the advanced class, most students would have very limited vocabularies and make considerable inflectional errors. Again, the cloze results were scored twice using two separate scoring procedures. Because of extreme vocabulary limitations, it was anticipated that most students would know no more than one possible word for each blank. This time, therefore, the exact scoring method was treated as the constant during both scorings, and only the exact word was accepted as correct. To get at inflectional considerations, during the first scoring everything had to be exact, including all inflections, in order to receive credit. During the second scoring, however, one-half credit was awarded if an improperly inflected exact word was supplied. By determining the correlations in the same manner as had been done with the class of advanced students, it was found that requiring correct inflections in order to receive any credit resulted in a reliability of .88. However, by awarding one-half credit for the

exact word incorrectly inflected, a reliability of .91 was achieved. Therefore, in this project the decision was made to award partial (one-half) credit in certain predetermined cases despite incorrect inflections.

Rules for Scoring the Cloze Test

Based on the findings presented above, a set of ten rules was established for scoring all cloze tests used in conjunction with this research. Although even higher reliabilities undoubtedly could have been achieved had even more detailed scoring procedures been used, it was felt essential to keep the rules relatively general and uncomplicated in order not to overwhelm and thereby discourage German teachers from using the findings of this research as a tool for assessing reading difficulty.

Rules established for scoring all cloze tests used in conjunction with this research include:

1. Assign two points to every cloze blank.

2. Accept as semantically correct the exact word deleted, a synonym of the exact word, or any contextually acceptable substitute that demonstrates comprehension.

3. Subtract one point for an incorrect noun-modifier or noun ending.

4. Subtract one point for an incorrect tense or tense form.

5. Subtract one point for an incorrect plural form.

6. Subtract one point for an article (definite or indefinite) that does not agree in gender, number, and case with the article that was deleted.

7. Unless contextually acceptable, award no credit if either an indefinite article is used to replace

a definite article or a definite article is used to replace an indefinite article.

8. Where a preposition is called for, accept as correct only the preposition that was deleted except in cases where two or more prepositions can be considered semantically correct. For Example, "Er sitzt an or hinter dem Tisch."

9. In all inflectional instances not described above, accept only an exact replacement.

10. Convert each raw cloze score into a percentage by dividing the total number of correct points by the total number of available points.

Rules for Scoring the Completion Test

Because the objective of establishing a cloze criterion score for reading based on information gain involves a comparison between cloze scores and completion-type objective test scores obtained from the same reading passage by matched students of German, it is essential that the rules for scoring the completion test also be as objective as possible. In developing rules for scoring the completion test in this research, guidelines laid down by Bormuth (1971) in his Study Two were heavily drawn from. Simplistically stated, the scoring procedure involved classifying responses as being either derivable or not derivable from the reading passage. Those responses that were derivable were scored correct; those that were not, were marked incorrect.

In determining which responses were derivable from the reading selection and thereby correct, all answers were examined for their relationship to the generic response, which was defined as either the word(s), phrase, or clause

replaced by the wh- words to form the questions. In answering the questions, it was possible for a correct answer to have either a formal or a semantic relationship to the generic response. If an answer was formally identical to the generic response, naturally it was scored correct. For example, if the generic response that had been deleted was "the man," the answer "the man" would be considered correct. If the answer was formally identical to only part of the generic response, only the word that was the grammatical head of the generic response was required in order for the answer to be scored correct. For example, with the generic response "the tall man," either "the tall man" or "the man" would be correct. If the generic response was a coordinated structure, only one of the coordinated structures was required. For example, with the generic response "golf and tennis," either "tennis and golf," "golf," or "tennis" would be correct.

In addition to formal considerations, semantic relationships between the generic response and the answer were also considered. Any answer with a correct semantic relationship to the generic response was counted correct. For example, if the correct generic response was "instruments," "tools" was also accepted as correct.

Finally, because the objective of the completion test was to determine how much the reader actually understood in order to make a valid point of comparison for the cloze test, those students taking the objective test were allowed

to answer in either German or English. Justification for this decision was based on the realization that if a reader knew the correct answer but could not think of the German word and therefore left the answer blank, maximum comprehension would not have been reflected and the ultimate validity of the findings would have been reduced. Therefore, English answers were accepted on the completion test although German responses were encouraged.

Statistical Analysis

In order to establish an objective and rational cloze criterion score based on information gain, it is necessary to determine some meaningful relationships by comparing the cloze test with an instrument that effectively measures information gain. When conventional completion tests are objectively constructed according to a set of predetermined, replicable rules, such tests can be used as the measure of information gain. To determine the exact relationship between the two test types, however, it is necessary to subject the data to a rigorous statistical analysis. Therefore, after having administered and scored both the cloze test and the completion test according to standards already outlined, all results obtained from the 105 matched pairs of readers were analyzed by means of regression analysis in order to predict information gain from the cloze scores and thereby establish an objective cloze criterion score. In the analysis, cloze scores were treated as the independent variable, while the objective test

scores were treated as the dependent variable. All analyses were performed either on an IBM 370 computer located at The Ohio State University (OSU) or on a Burroughs 6700 computer located at the United States Air Force Academy (USAFA). All analyses were accomplished using either the statistical program designated "Statistical Analysis System" (SAS) by Barr et al (1976) or the program designated "Statistical Package for the Social Sciences"(SPSS) by Nie et al (1975). All plotting presented in this research was done using the Calcomp Plotting System at USAFA.

CHAPTER IV

RESULTS AND DISCUSSION

The intent of Chapter IV is to analyze and discuss the findings of this research. To assure clarity and continuity, the chapter is divided into five sections: 1) a brief discussion of regression analysis, the primary statistical tool used to analyze the findings; 2) a look at the reliabilities of the two test types; 3) an examination of why and how the data were modified in preparation for analysis; 4) the actual analysis of the data with a discussion of the findings; and 5) answers to the three research questions.

Regression Analysis

When two separate measurements, say X and Y, have been made on a group of individuals, it is frequently desirable to study these measurements simultaneously in order to determine their relationship. To do so, several different procedures can be employed. For example, one can construct a simple scattergram, which employs two perpendicular axes, and plot each individual's scores. By examining the pattern of the plots, one can sometimes determine if and how the two measurements, generally

referred to as variables, are related. If more sophistication and precision are desired, the degree of relationship between the two measurements can be determined mathematically by computing a statistic referred to as the correlation coefficient (r). Should one be interested in using one of the measurements (the independent X variable) to explain or predict the other measurement (the dependent Y variable), a statistical procedure referred to as regression analysis is employed. The notion of regression analysis is closely related to that of correlation. In fact, the "r" used to indicate the correlation coefficient means "regression." Depending on the number of independent variables involved, regression analysis is traditionally divided into two broad categories, simple and multiple regression. In both instances the relationship between the independent variable(s) and the dependent variable can be either linear or nonlinear. Because the objective of this research is to examine only one independent variable, the cloze test, the topic of multiple regression will not be dealt with.

Simple Linear Regression

In simple regression a single independent variable X is used to explain or predict some dependent variable Y. To make this prediction, one must compute the best fitting regression curve of Y on X, i.e., that curve or line that accounts for the maximum amount of dependent variable variance. Of the various types of curves that

can be fitted to the data, the one most frequently used is the linear or straight line, probably because it is the simplest to compute. The basic equation used to express and compute simple linear regression is:

$$Y' = a + bX$$

In this equation X represents the various values of the independent variable, Y' is the predicted dependent variable, "a" is the intercept constant (that point at which the regression line intercepts the Y axis), and "b" is the regression coefficient (that number which identifies the slope of the regression line).

Curvilinear Regression

If for some reason a researcher is led to believe that the relationship between X and Y is something other than linear, possibly because of some underlying theory or simply by examining the data plots on a scattergram, it then becomes expedient to try and fit the data to some type of nonlinear mathematical model, such as a trigonometric or logarithmic function. For example, if a scattergram of the data revealed plots somewhat resembling those found in Figure 3, there would be a strong indication that the relationship between the variables resembled a logarithmic function. Logically, therefore, one would abandon fitting the data to a straight line and employ the following equation:

$$Y' = a + b \text{ LOGX}$$

which first causes the independent variable values to be transformed into their logarithmic function before being

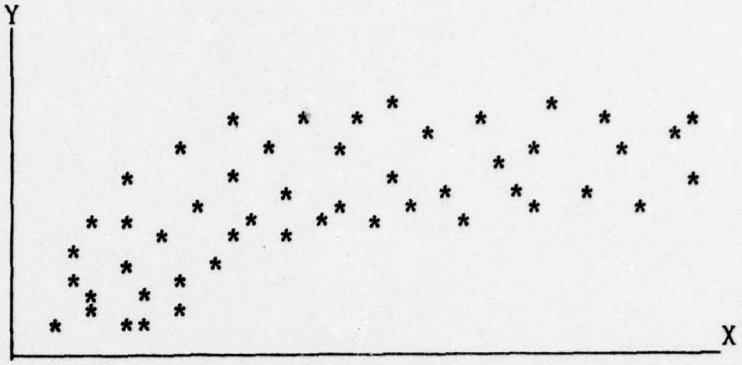


Figure 3. Simulated scattergram of nonlinear data points.

multiplied by the regression coefficient "b."

In addition to fitting data by using some type of trigonometric or logarithmic function, a different type of nonlinear regression referred to as polynomial regression analysis is often employed if some type of nonlinear trend in the data points is suspected. In performing polynomial regression, the independent variable is raised to some power, with the highest power representing the degree of the polynomial equation. As an example, the equation

$$Y' = a + b_1X + b_2X^2$$

is a second-degree polynomial, often referred to as a quadratic equation, because the independent variable X has been raised to the second power. The order of the polynomial equation indicates the maximum number of bends that can occur in the regression curve. For example, because the above equation is a second-degree polynomial containing two terms (X and X^2), it describes a single bend in the

regression curve and would provide an excellent fit for data similar to that found in Figure 4. Similarly, a third-degree polynomial equation containing three terms (X, X^2, and X^3) provides for two bends in the regression curve, and so on. According to Kerlinger and Pedhazur (1973), "the highest order that any given (polynomial) equation may take is equal to K − 1, where K is the number of distinct values in the independent variable" (p. 209).

Figure 4. Simulated scattergram of nonlinear data points.

Although the highest order polynomial equation would provide for K — 1 bends in the regression curve it generated, it would nevertheless have little predictive value and be of little use to the researcher. In this regard, Dunn and Clark (1974) indicate:

> It is possible to approximate any regression curve as closely as is desired by using a polynomial of sufficient high degree. However, in fitting a

> polynomial to a set of data, we are not just interested in obtaining a curve that lies very close to the sample data points. Rather, we are attempting to obtain a curve that approximates the population regression curve (line of population means). If our primary interest were in fitting the curve to the data points, we could actually fit a polynomial of such high degree that it would pass exactly through every data point. A straight line (polynomial of degree 1) can be passed through any two points; a quadratic (degree 2) can be passed through any three points; a cubic (degree 3) can be passed through any four points; an so on. Usually K, the degree of the fitted polynomial, is chosen to be 2 or 3; certainly K should be quite small compared with n, the sample size (p. 295).

Test Reliabilities

Prior to submitting the data to the computer for regression analysis, the reliabilities for both the cloze test and the completion test were computed to ensure that both test types were stable, consistent, and accurate. The reliability of each of the two test types was determined by first correlating all of the odd-numbered items in each test with all of the even-numbered items and then correcting (boosting) the resultant correlations by using the Spearman--Brown prophecy formula.

Although great pains had been taken in developing the completion-type test, as was expected, because of the cloze test's simplicity of construction, a higher reliability was realized on the cloze test. For the completion test a reliability of .84 was achieved, while on the cloze test a reliability of .95 was obtained. Because on the cloze exam three different test forms had been constructed and administered in order to counterbalance

any effects that might arise should the items in one form differ in difficulty from the other two forms, both the mean and reliability were computed for each of the three cloze forms. As shown in Table 6, there were indeed measurable differences in the three different forms of the cloze test--differences that could have affected the findings had only one form been used instead of three.

TABLE 6

CLOZE-TEST CORRELATIONS AND MEANS

Form	Correlation	Mean (%)
1	.96	50.6
2	.93	58.5
3	.96	50.8
All 3	.95	53.2

Data Modification

In preparing the data for analysis, all cloze scores were first adjusted in order to correct them for a statistical bias that had been pointed out by Bormuth. In discussing the regression models he had used in his Study One, Bormuth (1971) says the following:

> The regression models used in these studies do not take into account the error of measurement in the independent variable, the cloze scores. This error of measurement causes a student's observed cloze score to lie farther from the mean than its true value, the mean of the values that would be observed if he could be administered the test an infinite number of times. The effect of this bias would generally be to associate a gain score with an observed score that lay farther from the cloze mean than the true cloze score with which it is associated, thus stretching the curves at the extremes. Since this bias would influence the level at which a performance criterion was placed, it was removed by correcting the cloze scores to their true scores (p. 37).

In order to remove the above-mentioned bias, all cloze scores were transformed into what Bormuth refers to as their "true" scores by using the formula:

$$X_t = rX + (1 - r)\overline{X}$$

where X_t represents the student's true score, "r" the reliability of the test, X the student's observed score, and $\overline{X}$ the test mean.

To obtain the transformed score X_t for analysis, each of the 105 observed cloze test scores X were entered into the following equation:

$$X_t = .94999X + (1 - .94999)53.2104$$

Data Analysis

Simple Linear Regression Analysis

Having computed the reliabilities for both types of tests and further modified the independent variable, cloze test values in order to control for an error-of-measurement bias, the data were ready to be fitted to a regression model. First, however, to aid in determining exactly what type or types of regression models might be best to employ,

a scattergram of the data plots was obtained in order to make a preliminary visual analysis. As Figure 5 shows, at least the middle portion of the data plots appeared to be generally linear. As a beginning, therefore, it was decided to fit the data with a simple linear regression model.

Using the regression procedure contained in SAS, the following linear regression equation was derived:

$$Y' = 36.05553314 + .56959232X$$

which yielded the regression line found in Figure 6. That the linear regression model fit the data reasonably well and that there was a statistically significant relationship between the two test types was substantiated both by an analysis of variance (ANOVA), the findings of which are contained within Table 7, and by the fact that the correlation between the two variables was computed to be .73. This statistic squared ($r^2 = .53139$) yielded the coefficient of determination, which indicated that just over 53 percent of the variance detected in the completion-test scores could be accounted for by the cloze scores.

Polynomial Regression Analysis

Despite the fact that a simple linear regression model fit the data points rather well and accounted for a considerable amount of the variance, a plot of the regression line through the data points (see Figure 7) strongly suggested that using a more complex, curvilinear model

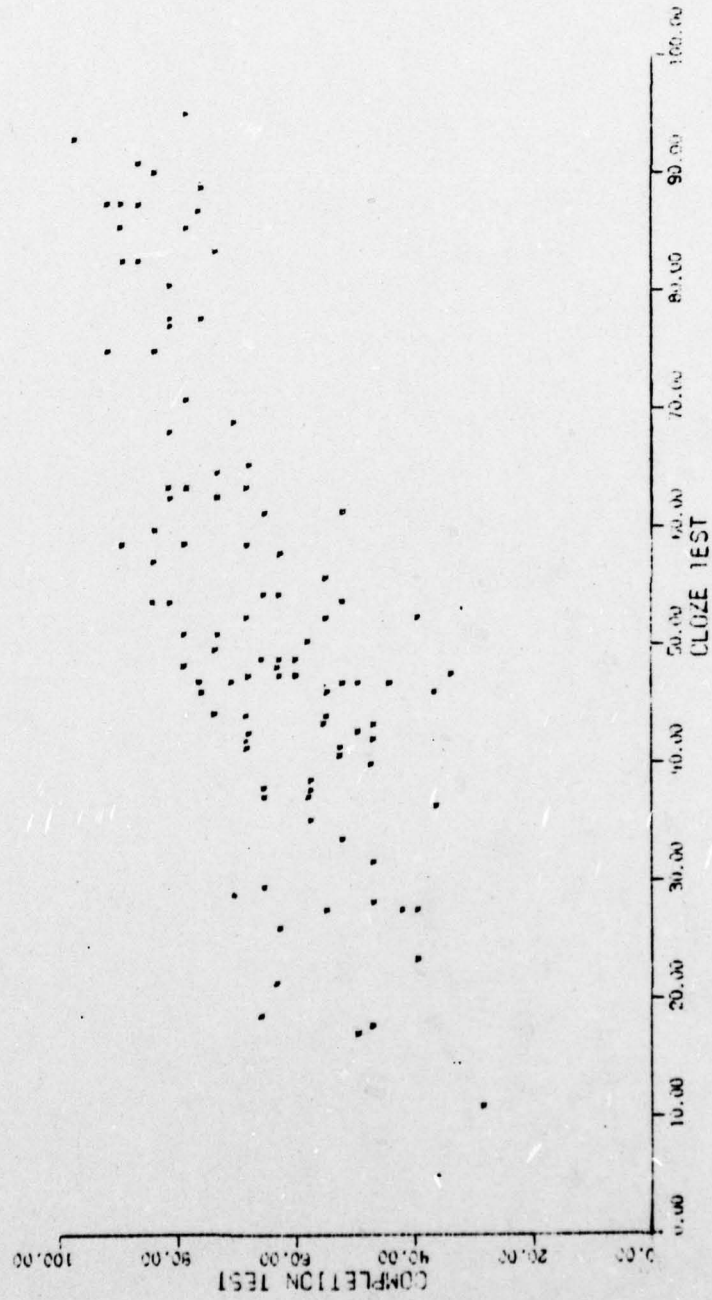


Figure 5. Scattergram of actual data points.

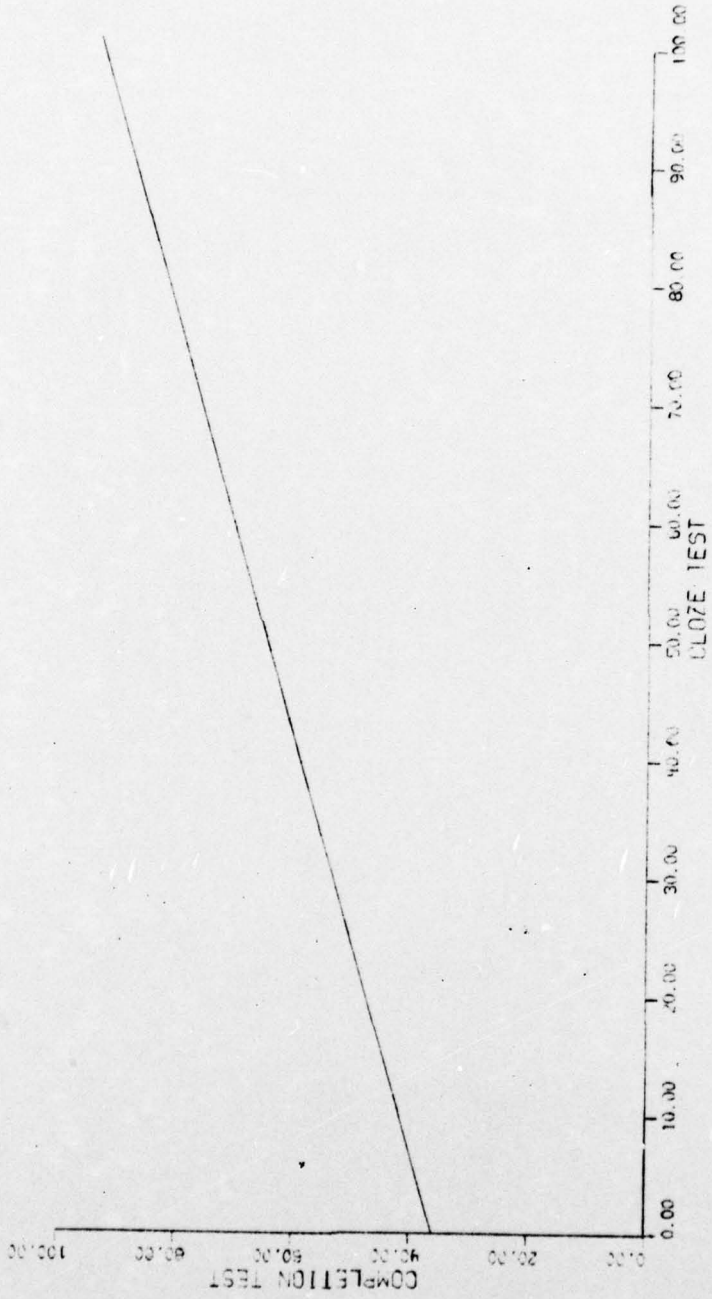


Figure 6. Plot of simple linear regression line.

TABLE 7

ANALYSIS OF VARIANCE OF SIMPLE LINEAR REGRESSION

Source	df	SS	MS	F
Regression	1	12618.96162	12618.96162	116.80*
Residual	103	11128.10085	108.03981	
Total	104	23747.06247		

*$p < .0001$

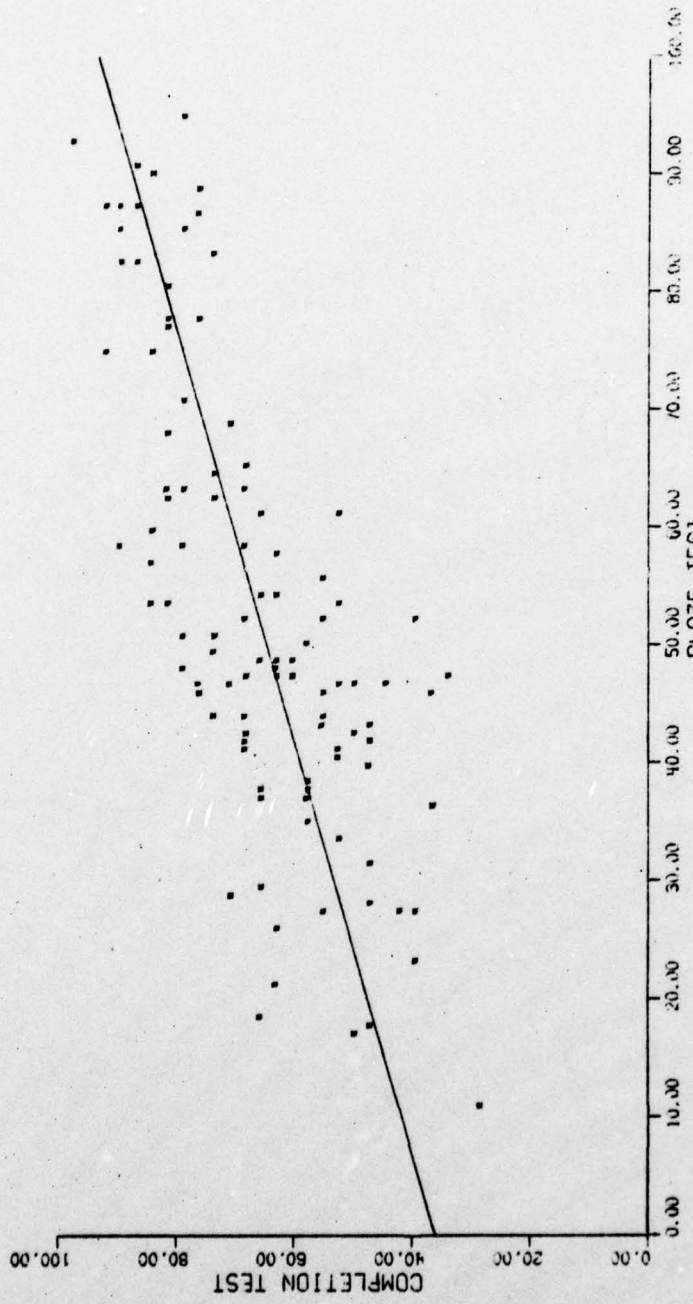


Figure 7. Simple linear regression line plotted through the specific data points.

would most likely account for additional variance and provide for a better fit.

Suspecting that the line of best fit would likely contain several bends, the decision was made to fit the data using a polynomial regression model. In choosing a statistical package for polynomial regression analysis, SAS was again chosen, primarily for two reasons: 1) The SAS package provides for two different types of F-tests of statistical significance: a) a sequential F-test (referred to by SAS as the Type I Sum of Squares), which adds the various powers (terms of the equation) of the independent variable sequentially one by one into the polynomial regression equation, and b) a partial F-test (referred to by SAS as the Type IV Sum of Squares), which treats each of the powers of the independent variable as though it were added to the equation last. According to Draper and Smith (1966), the partial F-test is important because it allows the researcher "to see the relative effects of each variable in excess of the others" (p. 72).

2) The SAS package, rather than simply supplying the F statistic, the significance of which generally must be determined by comparing it with a tabled value, also provides the computed confidence level corrected to the fourth decimal place.

In using the polynomial regression model, the independent variable, cloze-test scores were successively

raised in power ($b_1X + b_2X^2$, then $b_1X + b_2X^2 + b_3X^3$, etc.) from the second- up to the sixth-degree polynomial. At each of the separate powers, a regression analysis was run in order to determine which of the different polynomial regression equations would best fit the data. In analyzing the resultant sequential F-tests computed at each of the different powers (X^2 through X^6), totally insignificant F-values were obtained in every instance except when a fifth-degree polynomial was derived. At that level the computed F-value proved to be much more meaningful, resulting in a conficence level of just over .08 (see Table 8). As a confirmation of the significance of the fifth-degree polynomial, findings from the partial F-test (see Table 9) did, in fact, indicate that each of the five separate b-coefficient terms of the independent variable in the equation had a significant effect when added to the equation last. As Table 10 indicates, an analysis of the variance of the overall regression of the fifth-degree polynomial reflected an extremely significant F-value. Table 11, a summary of all the correlational relationships, reflects a high multiple correlation of more than .74 between all the components of the fifth-degree polynomial and the dependent variable, the completion-type comprehension test, and further indicates that such a regression model accounts for over 55 percent of the variance detected in the dependent variable.

TABLE 8

SEQUENTIAL F-TEST OF EACH POWER OF A FIFTH-DEGREE POLYNOMIAL REGRESSION

Source	df	SS	F-Value	Probability
X	1	12618.96162	118.89	.0001
X^2	1	0.18201	0.00	.9671
X^3	1	142.96747	1.35	.2486
X^4	1	163.86187	1.54	.2170
X^5	1	313.67703	2.96	.0887

TABLE 9

PARTIAL F-TEST OF EACH POWER OF A FIFTH-DEGREE POLYNOMIAL REGRESSION

Source	df	SS	F-Value	Probability
X	1	412.33658	3.89	.0515
X^2	1	392.41793	3.70	.0574
X^3	1	381.46511	3.59	.0609
X^4	1	351.89617	3.32	.0716
X^5	1	313.67703	2.96	.0887

TABLE 10

ANALYSIS OF VARIANCE OF FIFTH-DEGREE POLYNOMIAL REGRESSION

Source	df	SS	MS	F
Regression	5	13239.65002	2647.93000	24.95*
Residual	99	10507.41244	106.13547	
Total	104	23747.06246		

*$p < .0001$

TABLE 11

FIFTH-DEGREE POLYNOMIAL CORRELATIONAL SUMMARY TABLE

Variable	Multiple r	r-Square	r-Square Change	Simple r
X	.72897	.53139	.53139	.72897
X^2	.72897	.53140	.00001	.71453
X^3	.73309	.53742	.00602	.68328
X^4	.73778	.54432	.00690	.64900
X^5	.74668	.55753	.01321	.61664

As a result of the statistical significance of the fifth-degree polynomial, the decision was made to use the following equation to fit the data:

$$Y' = a + b_1X + b_2X^2 + b_3X^3 + b_4X^4 + b_5X^5$$

Substituting the coefficients derived from the fifth-degree polynomial regression analysis yielded the following equation:

$$Y' = -43.3892 + 11.59486X - .520828X^2 + .0109111X^3 - .00010399X^4 + .000000368X^5$$

which when plotted resulted in the regression curve found in Figure 8. In addition to accounting for more of the variance, Figure 9 shows that the fifth-degree polynomial curve, when compared with the line resulting from the linear equation (see Figure 7), appears to fit the data considerably better.

Stepwise Regression Analysis

Although the above findings indicated that the data points could be fitted well with a fifth-degree polynomial regression model, it was nevertheless decided to construct a second, different curvilinear model for purposes of comparison and confirmation. Instead of forcing all of the b-coefficient terms of the polynomial equation into the regression, however, as had been done in constructing the above fifth-degree polynomial model, it was decided to use a stepwise regression procedure. In stepwise regression the independent variable and its various nonlinear transformations that are selected for examination are

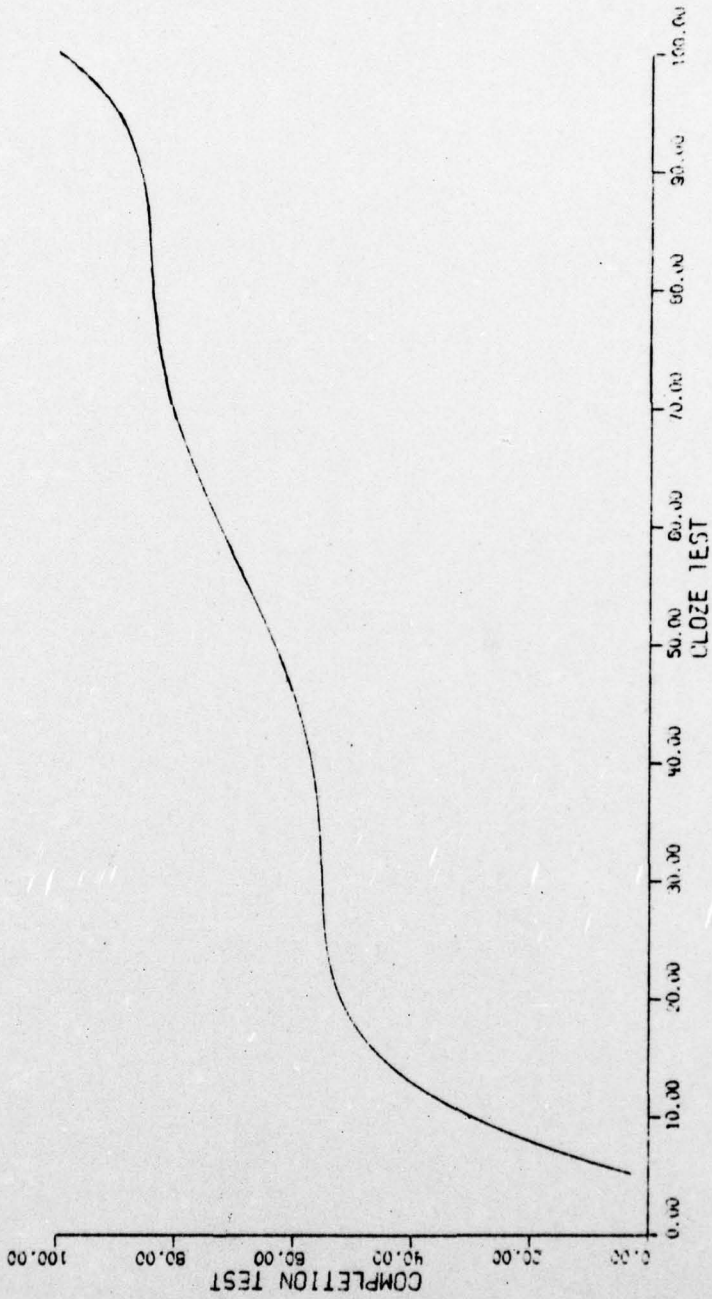


Figure 8. Plot of fifth-degree polynomial regression curve.

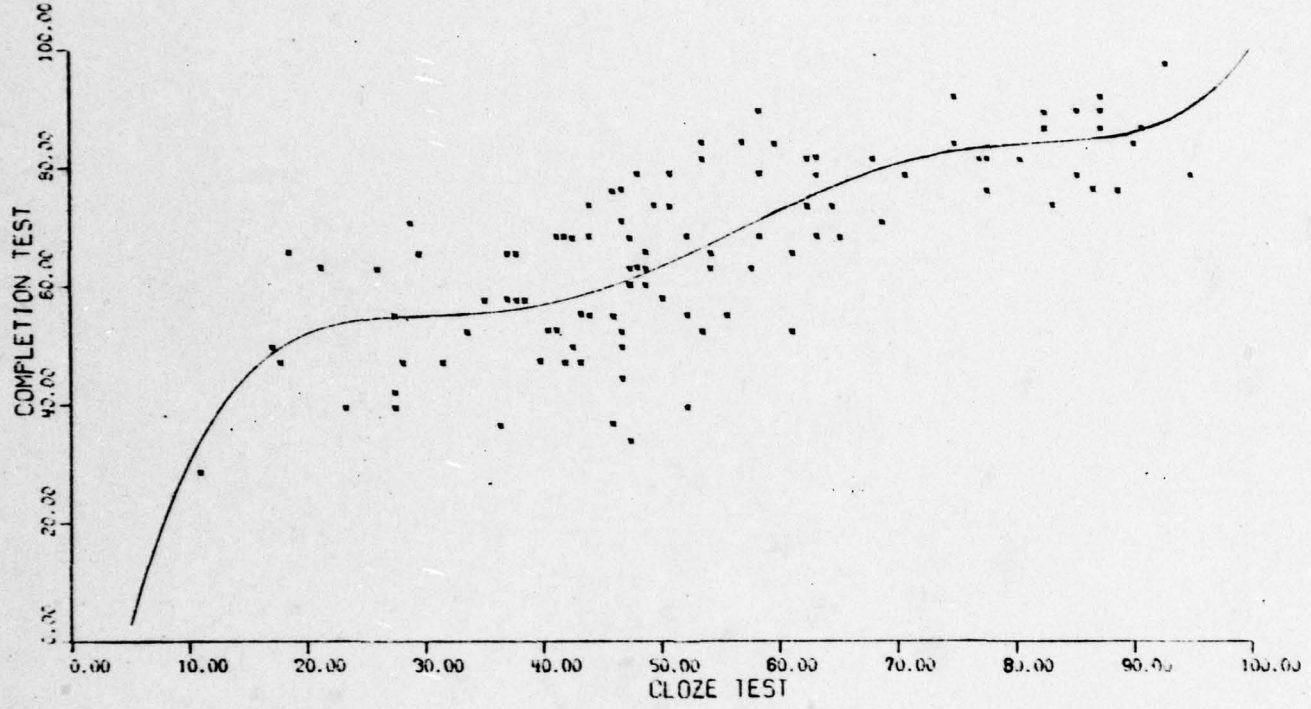


Figure 9. Fifth-degree polynomial regression curve plotted through the specific data points.

only entered into the equation if they meet certain statistical criteria. In addition, the order of their inclusion is determined not by the researcher, but by the respective contribution that each one makes in explaining variance. According to Nie et al (1975):

> The variable that explains the greatest amount of variance in the dependent variable will enter first; the variable that explains the greatest amount of variance in conjunction with the first will enter second, and so on. In other words, the variable that explains the greatest amount of variance unexplained by the variables already in the equation enters the equation at each step. And one or more of the variables may never be entered into the regression equation if the statistical criteria are not met (p. 345).

In deciding which nonlinear transformations of the independent-variable data to use in carrying out stepwise regression analysis, it seemed appropriate that the various power functions of the independent variable (X^2, X^3, etc.) should again be employed, at least up to the fifth power where high confidence levels had been derived with the first curvilinear regression model. In addition, however, it was decided to use two other types of nonlinear transformations, the logarithmic (LOG) function (see Figure ~~9~~ 10) and the arc tangent (ARC) function (see Figure ~~10~~ 11), both of which appeared might provide a good fit for the data and thereby account for additional variance.

Therefore, using the regression procedure contained in SPSS, the independent-variable scores (X) and six nonlinear transformations of those scores (X^2, X^3, X^3, X^5, LOGX, and ARCX) were submitted to a stepwise regression

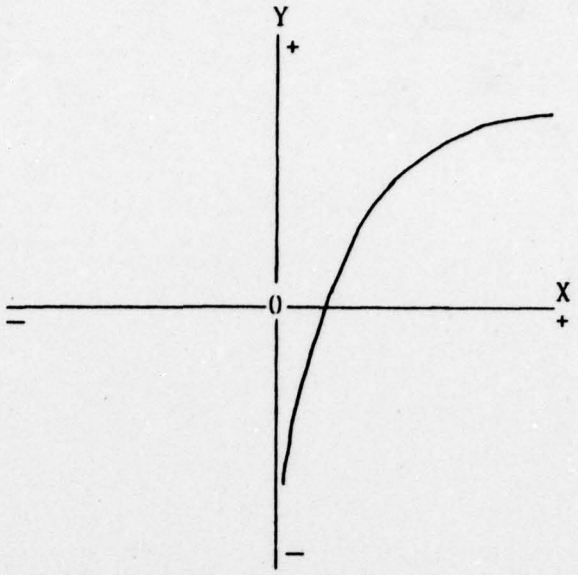


Figure 10. Simulated line produced by the logarithmic function.

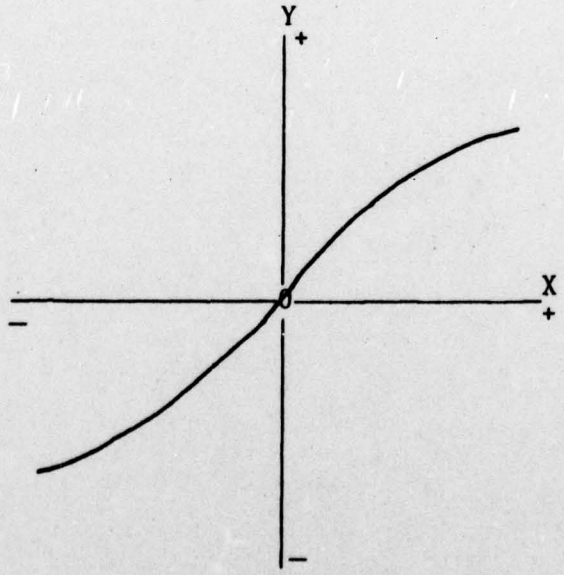


Figure 11. Simulated line produced by the arc tangent function.

analysis. As was expected, the first value to be selected was the nontransformed value X because, as had been discovered when employing the simple linear model, it alone could account for just over 53 percent of the variance in the dependent variable, completion-test scores. In order of their selection into the regression, those transformations of X that were able to account for statistically significant additional amounts of variance included: X^5, X^4, and ARCX.

As Table 12 indicates, an analysis of the variance of the overall regression containing the above four variables again reflects an extremely significant F-value. Table 13 summarizes the correlational relationships. Although the total amount of variance (r-square) that this model accounts for is also very high, a comparison with findings resulting from the fifth-degree polynomial model (see Table 11, p. 127) indicates that that model is slightly better, accounting for over one percent more of the variance detected in the dependent variable.

Despite the fact that the stepwise regression model was slightly less efficient than the fifth-degree polynomial model, the findings were nevertheless plotted as a means of comparison. Using the coefficients derived from the stepwise regression yielded the following equation:

$$Y' = .08866 - .37775X - .00033X^5 + .00365X^4 + 5.01913ARCX$$

which when plotted resulted in the curve found in Figure 13. As Figure 13 shows, both types of regression models

TABLE 12

ANALYSIS OF VARIANCE OF FOUR-TERM STEPWISE REGRESSION

Source	df	SS	MS	F
Regression	4	129.55450	32.38863	30.01278*
Residual	100	107.91612	1.07916	
Total	104	237.47062		

*p .0001

TABLE 13

FOUR-TERM STEPWISE CORRELATIONAL SUMMARY TABLE

Variable	Multiple r	r-Square	r-Square Change	Simple r
X	.72897	.53139	.53139	.72897
X^5	.72940	.53203	.00064	.61664
X^4	.73549	.54094	.00891	.64900
ARCX	.73862	.54556	.00462	.62072

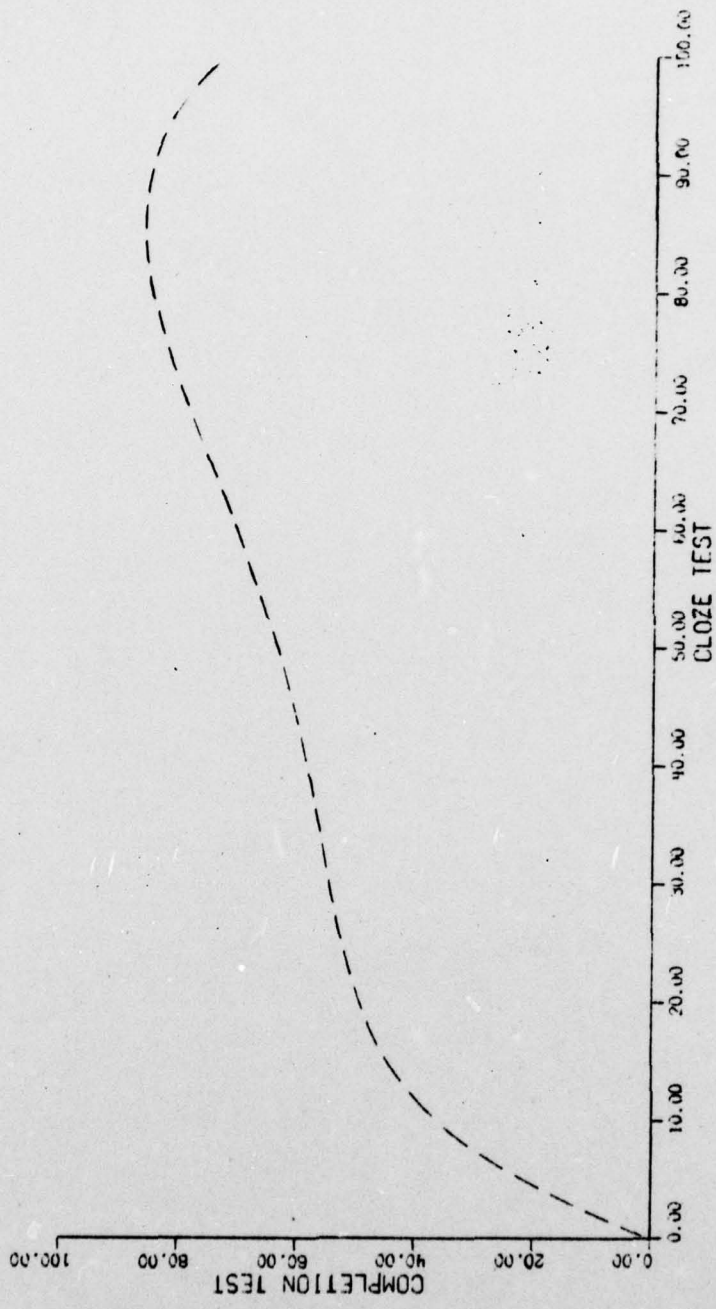


Figure 12. Plot of four-term stepwise regression curve.

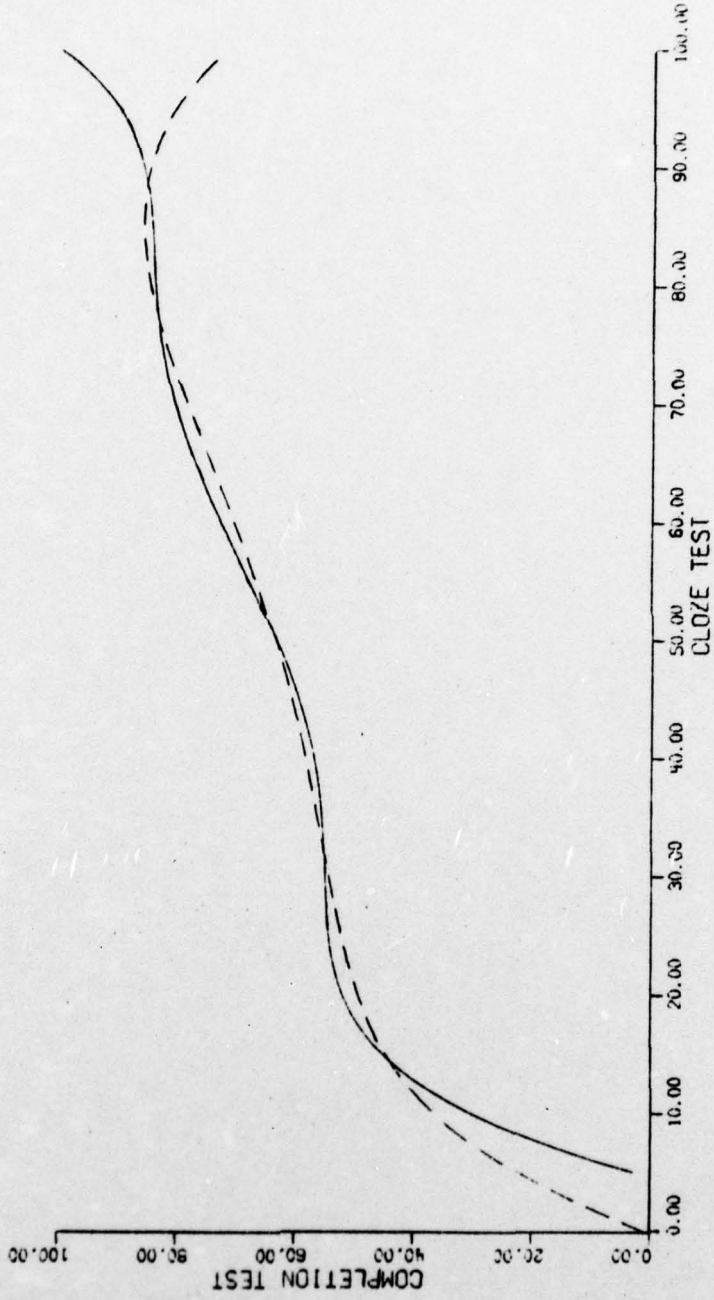


Figure 13. Plots of both nonlinear regression curves.

generate curves that are quite similar except at the extreme upper end where the arc tangent transformation of X apparently forces the regression curve of the stepwise model down. Figure 14 indicates that both curves fit the data points well, although, as might be expected, the fifth-degree polynomial curve appears to fit slightly better. Finally, Figure 15 depicts the regression lines of all three models (simple linear, fifth-degree polynomial, and four-term stepwise) plotted through the data points. In the central region all three models fit well. Again, however, overall, the fifth-degree polynomial model, which accounts for the largest amount of the variance, appears to provide the best fit. It is therefore this model that will be used in drawing the various conclusions of this research.

The Research Questions

Having analyzed the findings and identified the regression model that best fits the data to a regression curve for analysis, it is now appropriate to address the three research questions initially stated in Chapter I.

Question One

1. Is there a relationship between the cloze test and information gain, as measured by a specially constructed, short-answer completion test, when both test types are prepared from a single German prose reading selection, selectively administered to a carefully chosen group of college-level nonnative speakers of German, and then statistically analyzed?

Based on the findings of this research, there is a significant, measurable relationship between the cloze

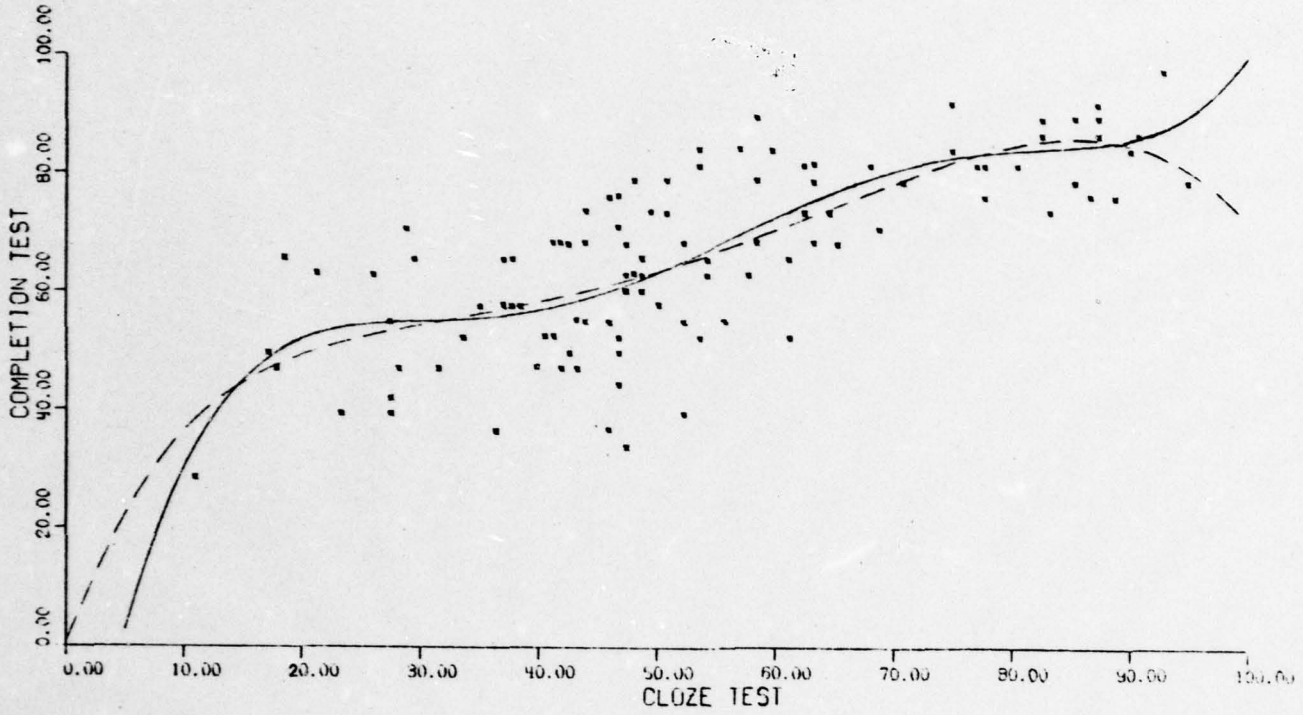


Figure 14. Both nonlinear regression curves plotted through the specific data points.

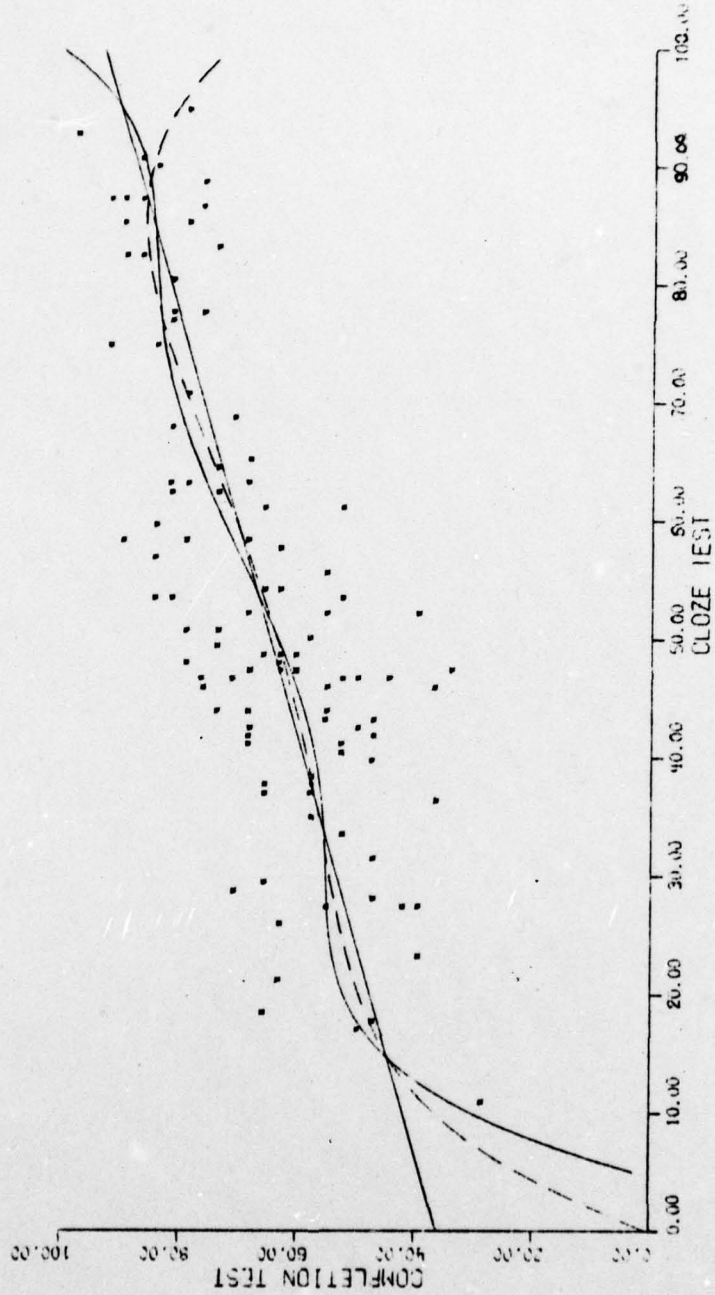


Figure 15. All three regression lines plotted through the specific data points.

test and information gain, as measured by the specially constructed, short-answer completion test. The simple correlation between the two tests is .73, with over 53 percent of the completion-test variance being accounted for by the cloze test. Additionally, by transforming the cloze-test values to their fifth power (X^5), a multiple correlation of .75 is achieved. Such a transformation results in nearly 56 percent of the completion-test variance being accounted for.

Question Two

2. If there is a relationship between the two test types, will it be nonlinear, as Bormuth's work in English suggests, when applied to nonnative speakers of German?

In confirmation of Bormuth's (1962 and 1971) cloze research in English, these findings also indicate that the relationship between the cloze test and the completion-type objective test is nonlinear when applied to nonnative speakers of German, a statement supported by the fact that a nonlinear (fifth-degree polynomial) regression model generated the regression curve that best fit the data. Of equal significance, however, is the fact that the plot of the fifth-degree polynomial curve (see Figure 8, p. 129) generally confirms Bormuth's contention that a cloze test discriminates over a broader range of reading abilities than does an objective test. Such a phenomenon is especially evident in the lower ranges (20 percent to 40 percent) of the cloze test where the slope of the regression curve is very close to zero.

The fact that the predicted completion test scores in this cloze range stand at slightly over 50 percent is not necessarily indicative that large amounts of significant information have been gained at this level as a result of reading. On the contrary, such a score on the completion test used in this study probably represents only spotty understanding of isolated information spread throughout the reading selection. This, because the test writing algorithm used to construct the completion test generated 38 questions, an extremely large amount for such a short passage. As a result, many of the questions are very easy, some probably as easy as some of the easiest cloze blanks, and as with the cloze test, can be answered without real understanding of the overall passage. As the regression curve shows, the cloze test does better discriminate, especially in the lower ranges, a fact strongly suggesting that an objective German cloze criterion score can be established that does indeed reflect that cloze percentage score where significant information begins to be won as a result of reading.

Question Three

3. If the relationship between the two test types is nonlinear, can a rational, objective cloze criterion score that is valid for any German prose reading selection be identified, which reflects that point on the cloze test at which nonnative readers of German begin gaining substantial information as a result of reading German prose?

Having shown that the cloze test and the completion-type objective test are highly related in a manner best

described by a fifth-degree polynomial regression curve, steps can now be taken to identify an objective cloze criterion score. Because of Bormuth's (1971) preliminary cloze research, which substantiated the psychometric assumption of regression identity that regression curves will not significantly change regardless of the number of reading passages used, the cloze criterion score to be established in this research can, in fact, be considered valid for any German prose reading selection.

As a first step in establishing the cloze criterion score, a visual analysis of the regression curve (see Figure 8) indicated that the score might be placed at about 35 percent, a point where the slope of the regression curve appears to begin to rise. In order to be more precise, however, the exact mathematical point of inflection was computed, i.e., that point along the lower end of the regression curve where the curvature reverses concavity. To make this computation, the fifth-degree polynomial regression equation was subjected to computer analysis. By taking the second derivative of the equation, setting it equal to zero, and solving for the appropriate cloze score, it was mathematically determined that the regression curve reverses concavity at exactly 30.1 percent on the cloze test. Technically, therefore, this cloze score (30 percent) can be considered to be the absolutely lowest point at which significant information may begin to be gained as measured by the cloze test.

In addition to establishing the mathematical point of inflection, which identifies the lowest point at which a cloze criterion score might be set, it was also felt that an upper point also would be of value by thus allowing the researcher to choose within the range of the two scores. Because the slope of the regression curve appears to approach linearity in its middle ranges, it was decided to calculate that point along the regression curve where its slope becomes equal to the slope of a straight line plotted through the data points. Using the b-coefficient (.5659232) that had been derived from the simple linear regression to determine that model's slope through the data, the fifth-degree polynomial equation was again entered into the computer. By taking the first derivative of the equation, setting it equal to .57, the b-coefficient of a straight line, and again solving for the appropriate cloze score, it was mathematically determined that the slope of the polynomial regression curve is equal to the curve of a straight line at 42.8 percent on the cloze test. Technically again, therefore, this score (43 percent) can be considered to be the absolute maximum at which significant information may begin to be gained as measured by the cloze test. Reader's scoring at or above this point definitely are gaining substantial amounts of information as a result of their reading.

Having established a range (30 percent through 43 percent) on the cloze test within which significant

information begins to be gained, the final step was simply to establish which specific score within that range should be selected as the cloze criterion score. This was accomplished by considering the standard error of estimate along the regression curve. Because the standard error was 10.3, it was decided to establish the cloze criterion score at 40 percent, approximately one standard error above the absolute low point of 30 percent. By setting the cloze criterion score at this point, one could be sure of being correct at least 68 percent of the time, especially if the cloze results were based on a class mean rather than an individual score.

As a result of the above reasons, 40 percent is the cloze score that has been chosen to identify that point at which significant information gain begins to occur when any German cloze test is scored in compliance with the ten rules presented in Chapter III (see p. 104).

CHAPTER V

SUMMARY, LIMITATIONS, IMPLICATIONS, AND RECOMMENDATIONS

Chapter V is divided into four parts: 1) a general summary of the objectives, procedures, and findings of this research; 2) an examination of the limitations; 3) the implications of this research for German teachers; and 4) suggestions for futher research.

Summary

Objective

Of the four language skills (speaking, listening, reading, and writing), many educators consider reading ultimately to be the most important because so much in the foreign language classroom centers around the student's ability to read. Despite this fact, however, teachers often unknowingly select reading materials far too difficult for their students' capabilities, primarily because the various measurement procedures presently available for determining the difficulty level of a reading selection are complicated, time consuming, and have all been shown to be generally subjective and unreliable. The overall goal of this research, therefore, was to

develop a simple, objective, valid, and reliable testing procedure to assist German teachers in choosing appropriately difficult reading selections for their students.

Procedure

To best accomplish the objective, it was decided to establish the relationship between the cloze test, a very simple device to construct, and information gain, as measured by a specially prepared, operationally replicable, completion-type objective test, and thereby establish a cloze criterion score by determining what percentage on German cloze tests, in general, represents that point at which nonnative readers begin to gain substantial information. To determine the relationship between the two test types, and thereby establish the cloze criterion score, both a cloze test and an operationally replicable completion test were constructed from the same German reading passage. In both instances, specific guidelines were laid out for test construction. 210 college-aged, nonnative students of German, who previously had been divided into 105 pairs matched according to German reading ability, were then selected and administered either the cloze test or the completion test so that each matched pair was represented by two scores, a cloze score and a completion-test score. After scoring the two types of tests according to specific, predetermined sets of rules, all findings were submitted to the computer for a regression analysis. In order to find the regression model

that would best fit the data and account for the largest amount of variance, several different types of linear and curvilinear regression models were experimented with. Ultimately, the data were fitted using a curve generated by a fifth-degree polynomial regression model.

Findings

In analyzing the results of the regression analysis, the two types of tests were found to be highly related. The multiple correlation between the five powers of the independent variable (cloze test) and the dependent variable (completion test) was .75 with the various powers of the cloze test accounting for nearly 56 percent of the variance found in the completion test. Because the relationship between the two tests was curvilinear, a phenomenon that had been expected based on previous cloze experimentation conducted in English by Bormuth, it was possible to establish a cloze criterion score that did, in fact, identify that cloze percentage point at which significant information began to be won as measured by the completion test. As a result, it was mathematically determined that 40 percent on the cloze test best described that level at which nonnative readers of German would begin to gain substantial information from reading German prose.

Limitations

In considering the findings of this research, the following two limitations should be considered:

The Sample

Because the sample of German students used in this study was limited to two central Ohio institutions of higher learning that had not been randomly selected from the larger population of all institutions of higher learning, and because the sample drawn from this limited population was not done so randomly, there exists the possibility that the reported findings are not completely generalizable to the larger population of all college-aged, nonnative students of German. No specific indication of such a circumstance was detected, however.

The Test-Writing Algorithm

A basic premise of this research is that in order for the derived cloze criterion score to be considered valid and generalizable, not only the cloze test, but also the objective test used to measure information gain must be operationally replicable. Otherwise, a variety of uncontrollable test-writer biases are introduced and something other than the actual difficulty level of the passage is measured, a circumstance that proportionately invalidates the findings. Thus, the need for the development of a test-writing algorithm to ensure objectivity.

Although the algorithm used in this research does, in fact, highly control the test-writing procedure, there are nevertheless two possible areas of concern: a) According to Rule 24, only case (C) and sentence(S) constituents may be replaced with an interrogative form to

generate a question. Such a rule excludes considering the verb, a form that is often key to sentence understanding and ideally should therefore be considered for deletion. Nevertheless, because of difficulties in generating semantically meaningful and structurally acceptable questions, the verb was purposely excluded from consideration. b) According to Rule 25, that constituent to be deleted is determined by random selection. This, of course, hinders complete operational replicability. Although different test writers using this algorithm would select identical sentences from a passage and analyze them in a similar manner, Rule 25 prevents having the exact same constituent replaced in all sentences.

Implications

In selecting student reading materials, most German teachers generally pay little attention to assessing carefully the difficulty level of the material they are examining. Primarily concerned with selecting the best subject matter, teachers generally determine the appropriateness of the difficulty level of the material either by following the recommendations of other teachers or seeing how the textbook writer himself has assessed the difficulty level of his work. Using such methods, of course, is highly subjective, unreliable, and likely invalid. No matter how appropriate or interesting reading material may be, if it is too difficult, the consequences will be disastrous.

Assuming that the findings of this research are valid and generalizable to all cloze tests constructed and scored according to guidelines laid out in Chapter III, there now exists, for the first time, a very simple procedure whereby teachers of college-aged, nonnative students of German can be assisted objectively in determining whether or not the reading materials they have selected for their students are, in fact, of the proper difficulty level. By choosing interesting reading material that has been objectively examined for difficulty, a serious source of frustration will have begun to be eliminated from the foreign language classroom.

Recommendations

The entire area of objectively assessing the appropriateness of foreign language reading materials with the cloze test is completely virgin territory. Consequently, there exists a host of varied topics to be investigated. Several that seem to be most directly related to this project include the following:

1) The findings of this research should be confirmed by replication using a sample that has been randomly drawn from a larger and therefore more representative population of colleges and universities.

2) Because the cloze criterion score established in this research can only be applied to college-aged students of German, there is a definite need to examine the topic of reading difficulty at the FLES, junior, and senior

high school levels where findings may be different because native-language reading skills often have not yet been fully developed.

3) To improve the validity of findings, additional research is needed in generating unbiased, operationally replicable objective tests with which to compare the cloze test.

4) This research needs to be replicated at all levels in other foreign languages. Of special interest are languages such as Arabic, Chinese, and Russian that have completely different writing systems.

5) Finally, the appropriateness of a reading selection is ultimately not solely dependent upon whether or not information is being gained. Other factors, such as motivation and style, are also important. Research should be conducted to identify, assess, and incorporate these other factors into the cloze criterion score.

APPENDIX A

READING SELECTIONS

Die Riesen und die Zwerge

Auf einem Schloss, hoch oben auf einem Berg, wohnte vor vielen Jahren ein Ritter mit seiner jungen Tochter. Der Ritter und seine Tochter waren beide Riesen. Deshalb waren beide, Vater und Kind, natürlich sehr gross und sehr stark. Trotz ihrer grossen Gestalt war die Tochter noch ein gutes Kind. Den ganzen Tag spielte sie gern.

Einst, an einem schönen Tag im Frühling, ging die Tochter spazieren. Wie immer, hatte sie es im Sinne, einen Spaziergang auf die nahe Spitze des Berges zu machen. Aber da war sie schon so oft gewesen. Weil sie einmal etwas Neues sehen wollte, ging sie ins Tal, wo sie noch nie gewesen war. Mit einigen grossen Riesenschritten war sie unten am Berge und auf dem Felde im Tal, wo die Bauern pflügten. Da stiess sie mit ihrem Fuss gegen etwas. Von ihrer grossen Höhe sah sie auf den Boden hinunter. Vor Freude über das, was sie sah, klatschte sie in die Hände. "Wie schön!" rief sie fröhlich. "Das muss ich meinem Vater zeigen!" Schnell kniete sie nieder, und mit der Hand fegte sie alles, was sie vor sich sah, in ihre grosse blaue Schürze. Fröhlich lief sie nach Hause. Vor Freude sprang sie leicht über den letzten steilen Felsen vor ihrem Schloss. "Vater, ach Vater, ich habe ein wunderschönes Spielzeug gefunden," rief sie und lief an den Tisch, wo ihr Vater beim Mittagsessen sass. "Ei, das muss etwas Schönes sein," sagte er. "Deine Augen leuchten ja vor Freude." "Sieh in meine Schürze! Kannst du sehen, was es ist?" fragte das Mädchen. "Es ist zu dunkel in der Schürze. Ich kann nichts sehen," sagte der Vater und lächelte. "Aber ich fühle, wie das Ding sich bewegt." Die Tochter öffnete nun ihre Schürze und stellte vorsichtig alles, was sie auf dem Felde gefunden hatte, vor dem Vater auf: einen Pflug, zwei Pferde und einen Bauern stellte sie auf den Tisch. Dann klatschte sie in die Hände und lachte laut über den komischen Zwerg, den Bauern, der voll Angst auf dem Tische umherlief. Auch der Vater lächelte über den komischen Anblick. Dann aber sagte er ernst zu seiner Tochter: "Mein Kind, das ist kein Spielzeug; gehe gleich wieder hinunter und bringe alles aufs Feld zurück." "Aber Vater, warum denn?" fragte das Mädchen ganz unglücklich als sie den Bauern und die Pferde in ihre Hände nahm. "Solch ein schönes lebendiges Spielzeig habe ich noch nie gehabt!" "Der Bauer ist kein Spielzeug, sage ich dir," sagte der Vater. "Er muss unten in dem Tal für uns arbeiten. Er muss pflügen, säen und ernten; sonst haben wir Riesen oben auf unseren Schlössern nichts zu essen und müssen hungern. Also trage ihn schnell dahin, wo du ihn gefunden hast." Die Tochter weinte, aber gehorsam packte sie den Pflug, die Pferde und den Bauern wieder in ihre Schürze und brachte sie auf das Feld zurück. "Wie schade,"

seufzte sie dann und trocknete ihre Tränen mit ihrer blauen Schürze. "Wie schade, dass mein Vater so streng ist."

Das Kamel

Einst schrieb eine berühmte Universität einen Preis aus für die beste schriftliche Arbeit über das Kamel. Unter den Kandidaten, die den Preis gewinnen wollten, waren ein Amerikaner, ein Engländer, ein Franzose und ein Deutscher. Jeder von ihnen schrieb einen Aufsatz über das Kamel, und jeder verfolgte dabei eine andere Methode.

Der praktische Engländer, der seinen Augen mehr traute als seinem Gehirn, packte seine Sachen und ging mit Badewanne und Teetasse in die Wüste Sahara. Dort lebte er auf einer Oase ganz wie in England. Jeden Tag badete er fleissig und trank viel Tee. Daneben spielte er Tennis und Cricket. Während seiner Freizeit studierte er das Kamel in der Wüste. Nach fünf Wochen hatte er genug Material für den Aufsatz und ging deshalb wieder nach England zurück. Dort schrieb er ohne grosse Probleme und in klaren kurzen Worten alles, was er in der Heimat der Kamele gesehen hatte.

In der Zwischenzeit hatte der Franzose seine Arbeit schon längst beendet. Er war in Paris geblieben und hatte einige Stunden vor den Kamelen im Zoo verbracht. Die Tiere mit ihrer phantastischen Gestalt hatten ihn sehr amüsiert. Sofort hatte er eine lustige Plauderei in seinem Kopf über das Kamel skizziert. Schnell war er in ein nahes Cafe gegangen, um die Skizze auszuarbeiten. Diese war voll lustiger und dramatischer Situationen. Wissenschaftlich war sie aber nicht.

Der Amerikaner hatte lange keine Zeit finden können, den Aufsatz zu schreiben. Er ist immer so furchtbar beschäftigt. Er ist Mitglied von zehn Klubs und Vorsitzender von zehn anderen. Endlich kommt der Tag, den er für die Fabrikation des Aufsatzes reserviert hatte. In die Wüste kann er nicht gut gehen. Den Zoo kann er auch nicht besuchen, weil er in Boston wohnt. Wo findet er nun so schnell wie möglich das nötige Material für seinen Aufsatz? Nur in der "Public Library" bekommt er die Information, die er braucht. Dort kann er alles, was je über das Kamel geschrieben worden ist, schnell überfliegen. Am nächsten Abend kommt er mit einem vollen Notizbuch nach Hause. Er setzt sich in den Schaukelstuhl und arrangiert sein Material. Sehr schnell stellt er mit Hilfe seiner Notizen eine Beschreibung des Kamels zusammen.

Der Deutsche hat viel mehr als alle andern zusammen geschrieben. Er hat sogar einen ganz dicken Band geschrieben. An demselben Tag, an dem er das Preisausschreiben der Universität gelesen hatte, schloss er sich in sein Studierzimmer ein. Niemand ausser seiner Frau durfte hereinkommen. Dort sass er im langen Schlafrock, mit einer noch

längeren Pfeife im Munde, und dachte Tag und Nacht über das Kamel. Von Zeit zu Zeit kam seine Frau auf den Zehenspitzen ins Zimmer, um ihm sein Essen zu bringen. Diese strenge Arbeit dauerte viele Wochen. Der arme Mann ass immer weniger und rauchte immer mehr. Nach zwei Monaten sah er wie ein Gespenst aus. Wenn man ihn sah, sass er immer in einer Rauchwolke vor seinem Schreibtisch. Mit gekrümmtem Rücken und gebeugtem Kopfe konnte man ihn wirklich kaum von einem Kamel unterscheiden. Langsam ging die Arbeit voran, denn der Gelehrte hatte es nicht leicht. Weil ein Deutscher alles immer gründlich macht, musste er das Kamel aus den Tiefen seiner eigenen Seele entwicklen. Und eine deutsche Seele ist sehr tief. Nach drei Monaten war das grosse Werk fertig, und der Gelehrte stiess einen sehr grossen Seufzer der Erleichterung aus. Mit triumphierender Freude schrieb er auf das Titelblatt des dicken Manuskriptes: "Das ideale Kamel, eine Philosophie des Unbewussten."

Wer, meinen Sie wohl, bekam den Preis?

APPENDIX B

INSTRUCTIONS AND TESTS

Name_____
(Please Print)

Instructor_____

Course Number_____

INTRODUCTION

Nothing is more frustrating for you, as a student, than to be required to read foreign language material with a difficulty level far above your own language capabilites--a phenomenon that can easily occur in the foreign language classroom. In order to solve this problem, you have been selected to participate in an educational research study designed to provide German instructors with empirical data that will enable them to accurately choose appropriate student reading selections.

You will be administered two short tests. Test I is to be administered today. Test II will be administered in the very near future. Please do your very best. The more conscientiously you participate, the more meaningful the findings will be to both instructors and future students of German.

INSTRUCTIONS

The attached test has been made by simply deleting every eighth word from a short story. Your task is to first read through the story to see what it is about and then to try to fill in each blank with the German word that makes most sense to you. You are to write only <u>one</u> word in each blank. If you are not sure of the word, do not be afraid to guess. Try to write an answer for every blank. Some blanks will be quite easy and some will be very difficult. To be most effective, do not spend too much time on any one blank. The answers to some items will be immediately apparent. By completing all of these items first and thereby providing additional context and meaning to the passage, you may increase your chances for answering some of the items you find to be more difficult. All of the blanks can be answered with ordinary German words. Wrong spelling will not be counted incorrect if the test grader can tell what word you intended. Please write neatly and remember, do <u>NOT</u> write more than one word in each blank. You should be able to complete this test in about 20-30 minutes. The test is not timed, however. When you are finished, raise your hand and your test material will be collected. If you have any questions, please ask them at this time.

(Do not turn the page
until told to do so.)

TEST I

Auf einem Schloß, hoch oben auf einem Berg, wohnte vor vielen Jahren ein Ritter mit seiner jungen Tochter. Der Ritter und _____ Tochter waren beide Riesen. Deshalb waren beide, _____ und Kind, natürlich sehr groß und sehr _____. Trotz ihrer großen Gestalt war die Tochter _____ ein gutes Kind. Den ganzen Tag spielte _____ gern.

Einst, an einem schönen Tag im _____, ging die Tochter spazieren. Wie immer, hatte _____ es im Sinne, einen Spaziergang auf die _____ Spitze des Berges zu machen. Aber da _____ sie schon so oft gewesen. Weil sie _____ etwas Neues sehen wollte, ging sie ins _____, wo sie noch nie gewesen war. Mit _____ großen Riesenschritten war sie unten am Berge _____ auf dem Felde im Tal, wo die _____ pflügten. Da stieß sie mit ihrem Fuß _____ etwas. Von ihrer großen Höhe sah sie _____ den Boden hinunter. Vor Freude über das, _____ sie sah, klatschte sie in die Hände. "_____ schön!" rief sie fröhlich. "Das muß ich _____ Vater zeigen!" Schnell kniete sie nieder, und _____ der Hand fegte sie alles, was sie _____ sich sah, in ihre große blaue Schürze. _____ lief sie nach Hause. Vor Freude sprang _____ leicht über den letzten steilen Felsen vor _____ Schloß. "Vater, ach Vater, ich habe ein _____ Spielzeug gefunden," rief sie und lief an _____ Tisch, wo ihr Vater beim Mittagessen saß. "_____, das muß etwas Schönes sein," sagte er. "_____ Augen leuchten ja vor Freude." "Sieh in _____ Schürze! Kannst du sehen, was es ist?" _____ das Mädchen. "Es ist zu dunkel in _____ Schürze. Ich kann nichts sehen," sagte der _____ und lächelte. "Aber ich fühle wie das _____ sich bewegt." Die Tochter öffnete nun ihre _____ und stellte vorsichtig alles, was sie auf _____ Felde gefunden hatte, vor dem Vater auf: _____ Pflug, zwei Pferde und einen Bauern stellte _____ auf den Tisch. Dann klatschte sie in _____ Hände und lachte laut über

den komischen _____, den Bauern, der voll Angst auf dem _____ umherlief. Auch der Vater lächelte über den _____ Anblick. Dann aber sagte er ernst zu _____ Tochter: "Mein Kind, das ist kein Spielzeug; _____ gleich wieder hinunter und bringe alles aufs _____ zurück." "Aber Vater, warum denn?" fragte das _____ ganz unglücklich als sie den Bauern und _____ Pferde in ihre Hände nahm. "Solch ein _____ lebendiges Spielzeug habe ich noch nie gehabt!" "_____ Bauer ist kein Spielzeug, sage ich dir," _____ der Vater. "Er muß unten in dem _____ für uns arbeiten. Er muß pflügen, säen _____ ernten; sonst haben wir Riesen oben auf _____ Schlössern nichts zu essen und müssen hungern. _____ trage ihn schnell dahin, wo du ihn _____ hast." Die Tochter weinte, aber gehorsam packte _____ den Pflug, die Pferde und den Bauern _____ in ihre Schürze und brachte sie auf _____ Feld zurück. "Wie schade," seufzte sie dann _____ trocknete ihre Tränen mit ihrer blauen Schürze. "Wie schade, daß mein Vater so streng ist!"

Name_____

Instructor_____

Course Number_____

INSTRUCTIONS

The attached test has been constructed in a manner similar to the last test you were administered, i.e., every eighth word has simply been deleted from a short story. Your task is to first read through the story to see what it is about and then to try to fill in each blank with the German word that makes most sense to you. You are to write only one word in each blank. If you are not sure of the word, do not be afraid to guess. Try to write an answer for every blank. Some blanks will be quite easy and some will be very difficult. To be most effective, do not spend too much time on any one blank. The answers to some items will be immediately apparent. By completing all of these items first and thereby providing additional context and meaning to the passage, you may increase your chances for answering some of the items you find to be more difficult. All of the blanks can be answered with ordinary German words. Wrong spelling will not be counted incorrect if the test grader can tell what word you intended. Please write neatly and remember, do NOT write more than one word in each blank. If you have any questions, please raise your hand. Otherwise, you may begin work now.

TEST IIA (Form 1)

Einst schrieb eine berühmte Universität einen Preis aus für die beste schriftliche Arbeit über das Kamel. Unter _____ Kandidaten, die den Preis gewinnen wollten, waren _____ Amerikaner, ein Engländer, ein Franzose und ein _____. Jeder von ihnen schrieb einen Aufsatz über _____ Kamel, und jeder verfolgte dabei eine andere _____..

Der praktische Engländer, der seinen Augen mehr _____ als seinem Gehirn, packte seine Sachen und _____ mit Badewanne und Teetasse in die Wüste _____. Dort lebte er auf einer Oase ganz _____ in England. Jeden Tag badete er fleißig _____ trank viel Tee. Daneben spielte er Tennis _____ Cricket. Während seiner Freizeit studierte er das _____ in der Wüste. Nach fünf Wochen hatte _____ genug Material für den Aufsatz und ging _____ wieder nach England zurück. Dort schrieb er _____ große Probleme und in klaren kurzen Worten _____, was er in der Heimat der Kamele _____ hatte.

In der Zwischenzeit hatte der Franzose _____ Arbeit schon längst beendet. Er war in _____ geblieben und hatte einige Stunden vor den _____ im Zoo verbracht. Die Tiere mit ihrer _____ Gestalt hatten ihn sehr amüsiert. Sofort hatte _____ eine lustige Plauderei in seinem Kopf über _____ Kamel skizziert. Schnell war er in ein _____ Cafe´gegangen, um die Skizze auszuarbeiten. Diese _____ voll lustiger und dramatischer Situationen. Wissenschaftlich war _____ aber nicht.

Der Amerikaner hatte lange keine _____ finden können, den Aufsatz zu schreiben. Er _____ immer so furchtbar beschäftigt. Er ist Mitglied _____ zehn Klubs und Vorsitzender von zehn anderen. _____ kommt der Tag, den er für die _____ des Aufsatzes reserviert hatte. In die Wüste _____ er nicht gut gehen. Den Zoo kann _____ auch nicht besuchen, weil er in Boston _____. Wo findet er nun so schnell wie _____ das nötige Material für seinen Aufsatz? Nur _____ der "Public Library" bekommt er die

Information, _____ er braucht. Dort kann er alles, was _____ über das Kamel geschrieben worden ist, schnell _____. Am nächsten Abend kommt er mit einem _____ Notizbuch nach Hause. Er setzt sich in _____ Schaukelstuhl und arrangiert sein Material. Sehr schnell _____ er mit Hilfe seiner Notizen eine Beschreibung _____ Kamels zusammen.

Der Deutsche hat viel mehr _____ alle andern zusammen geschrieben. Er hat sogar _____ ganz dicken Band geschrieben. An demselben Tag, _____ dem er das Preisausschreiben der Universität gelesen _____, schloß er sich in sein Studierzimmer ein. _____ außer seiner Frau durfte hereinkommen. Dort saß _____ im langen Schlafrock, mit einer noch längeren _____ im Munde, und dachte Tag und Nacht _____ das Kamel. Von Zeit zu Zeit kam _____ Frau auf den Zehenspitzen ins Zimmer, um _____ sein Essen zu bringen. Diese strenge Arbeit _____ viele Wochen. Der arme Mann aß immer _____ und rauchte immer mehr. Nach zwei Monaten _____ er wie ein Gespenst aus. Wenn man _____ sah, saß er immer in einer Rauchwolke _____ seinem Schreibtisch. Mit gekrümmtem Rücken und gebeugtem _____ konnte man ihn wirklich kaum von einem _____ unterscheiden. Langsam ging die Arbeit voran, denn _____ Gelehrte hatte es nicht leicht. Weil ein _____ alles immer gründlich macht, mußte er das _____ aus den Tiefen seiner eigenen Seele entwickeln. _____ eine deutsche Seele ist sehr tief. Nach _____ Monaten war das große Werk fertig, und _____ Gelehrte stieß einen sehr großen Seufzer der _____ aus. Mit triumphierender Freude schrieb er auf _____ Titelblatt des dicken Manuskriptes: "Das ideale Kamel, _____ Philosophie des Unbewußten."

Wer, meinen Sie wohl, bekam den Preis?

TEST IIA (Form 2)

Einst schrieb eine berühmte Universität einen Preis aus für die beste schriftliche Arbeit über das Kamel. Unter den Kandidaten, die _____ Preis gewinnen wollten, waren ein Amerikaner, ein _____, ein Franzose und ein Deutscher. Jeder von _____ schrieb einen Aufsatz über das Kamel, und _____ verfolgte dabei eine andere Methode.

Der praktische _____, der seinen Augen mehr traute als seinem _____, packte seine Sachen und ging mit Badewanne _____ Teetasse in die Wüste Sahara. Dort lebte _____ auf einer Oase ganz wie in England. _____ Tag badete er fleißig und trank viel _____. Daneben spielte er Tennis und Cricket. Während _____ Freizeit studierte er das Kamel in der _____. Nach fünf Wochen hatte er genug Material _____ den Aufsatz und ging deshalb wieder nach _____ zurück. Dort schrieb er ohne große Probleme _____ in klaren kurzen Worten alles, was er _____ der Heimat der Kamele gesehen hatte.

In _____ Zwischenzeit hatte der Franzose seine Arbeit schon _____ beendet. Er war in Paris geblieben und _____ einige Stunden vor den Kamelen im Zoo _____. Die Tiere mit ihrer phantastischen Gestalt hatten _____ sehr amüsiert. Sofort hatte er eine lustige _____ in seinem Kopf über das Kamel skizziert. _____ war er in ein nahes Cafe' gegangen, _____ die Skizze auszuarbeiten. Diese war voll lustiger _____ dramatischer Situationen. Wissenschaftlich war sie aber nicht.

_____ Amerikaner hatte lange keine Zeit finden können, _____ Aufsatz zu schreiben. Er ist immer so _____ beschäftigt. Er ist Mitglied von zehn Klubs _____ Vorsitzender von zehn anderen. Endlich kommt der _____, den er für die Fabrikation des Aufsatzes _____ hatte. In die Wüste kann er nicht _____ gehen. Den Zoo kann er auch nicht _____, weil er in Boston wohnt. Wo findet _____ nun so schnell wie möglich das nötige _____ für seinen Aufsatz? Nur in der "Public _____" bekommt er die

Information, die er braucht. _____ kann er alles, was je über das _____ geschrieben worden ist, schnell überfliegen. Am nächsten _____ kommt er mit einem vollen Notizbuch nach _____. Er setzt sich in den Schaukelstuhl und _____ sein Material. Sehr schnell stellt er mit _____ seiner Notizen eine Beschreibung des Kamels zusammen.

_____ Deutsche hat viel mehr als alle andern _____ geschrieben. Er hat sogar einen ganz dicken _____ geschrieben. An demselben Tag, an dem er _____ Preisausschreiben der Universität gelesen hatte, schloß er _____ in sein Studierzimmer ein. Niemand außer seiner _____ durfte hereinkommen. Dort saß er im langen _____, mit einer noch längeren Pfeife im Munde, _____ dachte Tag und Nacht über das Kamel. _____ Zeit zu Zeit kam seine Frau auf _____ Zehenspitzen ins Zimmer, um ihm sein Essen _____ bringen. Diese strenge Arbeit dauerte viele Wochen. _____ arme Mann aß immer weniger und rauchte _____ mehr. Nach zwei Monaten sah er wie _____ Gespenst aus. Wenn man ihn sah, saß _____ immer in einer Rauchwolke vor seinem Schreibtisch. _____ gekrümmtem Rücken und gebeugtem Kopfe konnte man _____ wirklich kaum von einem Kamel unterscheiden. Langsam _____ die Arbeit voran, denn der Gelehrte hätte _____ nicht leicht. Weil ein Deutscher alles immer _____ macht, mußte er das Kamel aus den _____ seiner eigenen Seele entwickeln. Und eine Deutsche _____ ist sehr tief. Nach drei Monaten war _____ große Werk fertig, und der Gelehrte stieß _____ sehr großen Seufzer der Erleichterung aus. Mit _____ Freude schrieb er auf das Titelblatt des _____ Manuskriptes: "Das ideale Kamel, eine Philosophie des _____."

Wer, meinen Sie wohl, bekam den Preis?

TEST IIA (Form 3)

Einst schrieb eine berühmte Universität einen Preis aus für die beste schriftliche Arbeit über das Kamel. Unter den Kandidaten, _____ den Preis gewinnen wollten, waren ein Amerikaner, _____ Engländer, ein Franzose und ein Deutscher. Jeder _____ ihnen schrieb einen Aufsatz über das Kamel, _____ jeder verfolgte dabei eine andere Methode.

Der _____ Engländer, der seinen Augen mehr traute als _____ Gehirn, packte seine Sachen und ging mit _____ und Teetasse in die Wüste Sahara. Dort _____ er auf einer Oase ganz wie in _____. Jeden Tag badete er fleißig und trank _____ Tee. Daneben spielte er Tennis und Cricket. _____ seiner Freizeit studierte er das Kamel in _____ Wüste. Nach fünf Wochen hatte er genug _____ für den Aufsatz und ging deshalb wieder _____ England zurück. Dort schrieb er ohne große _____ und in klaren kurzen Worten alles, was _____ in der Heimat der Kamele gesehen hatte. _____ der Zwischenzeit hatte der Franzose seine Arbeit _____ längst beendet. Er war in Paris geblieben _____ hatte einige Stunden vor den Kamelen im _____ verbracht. Die Tiere mit ihrer phantastischen Gestalt _____ ihn sehr amüsiert. Sofort hatte er eine _____ Plauderei in seinem Kopf über das Kamel _____. Schnell war er in ein nahes Cafe' _____, um die Skizze auszuarbeiten. Diese war voll _____ und dramatischer Situationen. Wissenschaftlich war sie aber _____.

Der Amerikaner hatte lange keine Zeit finden _____, den Aufsatz zu schreiben. Er ist immer _____ furchtbar beschäftigt. Er ist Mitglied von zehn _____ und Vorsitzender von zehn anderen. Endlich kommt _____ Tag, den er für die Fabrikation des _____ reserviert hatte. In die Wüste kann er _____ gut gehen. Den Zoo kann er auch _____ besuchen, weil er in Boston wohnt. Wo _____ er nun so schnell wie möglich das _____ Material für seinen Aufsatz? Nur in der "_____ Library" bekommt er die Information, die er _____. Dort kann er alles, was je

über _____ Kamel geschrieben worden ist, schnell überfliegen. Am _____ Abend kommt er mit einem vollen Notizbuch _____ Hause. Er setzt sich in den Schaukelstuhl _____ arrangiert sein Material. Sehr schnell stellt er _____ Hilfe seiner Notizen eine Beschreibung des Kamels _____.

Der Deutsche hat viel mehr als alle _____ zusammen geschrieben. Er hat sogar einen ganz _____ Band geschrieben. An demselben Tag, an dem _____ das Preisausschreiben der Universität gelesen hatte, schloß _____ sich in sein Studierzimmer ein. Niemand außer _____ Frau durfte hereinkommen. Dort saß er im _____ Schlafrock, mit einer noch längeren Pfeife im _____, und dachte Tag und Nacht über das _____. Von Zeit zu Zeit kam seine Frau _____ den Zehenspitzen ins Zimmer, um ihm sein _____ zu bringen. Diese strenge Arbeit dauerte viele _____. Der arme Mann aß immer weniger und _____ immer mehr. Nach zwei Monaten sah er _____ ein Gespenst aus. Wenn man ihn sah, _____ er immer in einer Rauchwolke vor seinem _____. Mit gekrümmtem Rücken und gebeugtem Kopfe konnte _____ ihn wirklich kaum von einem Kamel unterscheiden. _____ ging die Arbeit voran, denn der Gelehrte _____ es nicht leicht. Weil ein Deutscher alles _____ gründlich macht, mußte er das Kamel aus _____ Tiefen seiner eigenen Seele entwickeln. Und eine _____ Seele ist sehr tief. Nach drei Monaten _____ das große Werk fertig, und der Gelehrte _____ einen sehr großen Seufzer der Erleichterung aus. _____ triumphierender Freude schrieb er auf das Titelblatt _____ dicken Manuskriptes: "Das ideale Kamel, eine Philosophie _____ Unbewußten."

Wer, meinen Sie wohl, bekam den Preis?

Name_____

Instructor_____

Course Number_____

INSTRUCTIONS

Attached is a short story in German. You are to read this story until you feel you understand it as well as you personally are capable. To achieve this, it is suggested that you read the story carefully through at least two or three times. When you are satisfied that you have achieved your maximum understanding level of the story, raise your hand. The story will then be collected and you will be given a 38-question completion test to measure your comprehension of the passage. The test questions are all short and can be answered in one sentence or less. Some questions are rather general and therefore quite easy, and some require understanding and recollection of rather specific information. Remember, you will NOT have access to the story while you are taking the test. It is therefore extremely important that you read the story very carefully. If you have any questions, please raise your hand. Otherwise, begin reading the attached story now.

Name_____
(Please Print)

Instructor_____

Course Number_____

TEST IIB

This test is intended to measure your understanding of the short story you have just read. It consists of 38 short questions that can be answered in one sentence or less. Some questions are very easy, and some require an understanding and recollection of rather specific information. Because the test is intended to measure your understanding of the story, you may answer the questions in either German or English. If possible, answer them in German. You will not be penalized for incorrect grammatical constructions or spelling. Nevertheless, on any given question in which you feel you can express yourself more fully or clearly in English, please do so. Remember, the objective of the test is to measure your understanding of the story. Please write clearly and neatly. When you have finished the test, raise your hand and your test booklet will be collected.

1. Was schrieb einst eine berühmte Universität aus?_____

2. Wer schrieb einen Aufsatz über das Kamel?_____

3. Wo studierte der Engländer das Kamel während seiner Freizeit?

4. Was hatte der Franzose in seinem Kopf sofort skizziert?_____

5. Was alles kann der Amerikaner in der "Public Library" schnell überfliegen?

6. Wer dachte Tag und Nacht über das Kamel?_____

7. Wovon konnte man den Deutschen, mit gekrümmtem Rücken und gebeugtem Kopfe, wirklich kaum unterscheiden?

8. Wie mußte der Deutsche das Kamel entwickeln, weil er alles immer gründlich macht?

9. Wer schrieb mit triumphierender Freude auf das Titelblatt des dicken Manuskriptes: Das ideele Kamel, eine Philosophie des Unbewußten?

10. Wo lebte der Engländer in der Wüste ganz wie in England?_____

_____

11. Wohin ging der Engländer wieder zurück, weil er genug Material für den Aufsatz hatte?

_____

12. Wie lange hatte der Franzose vor den Kamelen im Zoo verbracht?

_____

13. Wer kann den Zoo nicht besuchen, weil er in Boston wohnt?____

_____

14. Wer badete jeden Tag fleißig?_____

_____

15. Wohin ging der Engländer mit Badewanne und Teetasse?_____

_____

16. Wer hatte in der Zwischenzeit seine Arbeit schon längst beendet?

_____

17. Was spielte der Engländer daneben?_____

_____

18. Wer außer der Frau des Deutschen durfte ins Studierzimmer hereinkommen?

_____

19. Womit stellte der Amerikaner sehr schnell eine Beschreibung des Kamels zusammen?

_____

20. Wo war der Franzose geblieben?_____

_____

21. An welchem Tag schloß der Deutsche sich in sein Studierzimmer ein?

_____

22. Was bekommt der Amerikaner nur in der "Public Library"?_____

_____

23. Worin setzt sich der Amerikaner?_____

_____

24. Warum kam die Frau des Deutschen auf den Zehenspitzen ins Zimmer von Zeit zu Zeit?

_____

25. Womit kommt der Amerikaner am nächsten Abend nach Hause?_____

_____

26. Was alles schriet der Engländer in England ohne große Probleme und in klaren kurzen Worten?

_____

27. Wer rauchte immer mehr?_____

_____

28. Was für ein Tag kommt endlich?_____

_____

29. Wovor saß der Deutsche immer in einer Rauchwolke, wenn man ihn sah?

_____

30. In was für ein Cafe'war der Franzose schnell gegangen, um die Skizze auszuarbeiten?

_____

31. Unter was für Kandidaten waren ein Amerikaner, ein Engländer, ein Franzose und ein Deutscher?

_____

32. Was zu tun, hatte der Amerikaner lange keine Zeit finden können?

_____

33. Wen hatten die Tiere mit ihrer phantastischen Gestalt sehr amüsiert?

_____

34. Wer ist Mitglied von zehn Klubs?_____

_____

35. Was war voll lustiger und dramatischer Situationen?_____

_____

36. Wer saß in seinem Studierzimmer mit einer noch längeren Pfeife im Munde?

_____

37. Was für ein Engländer packte seine Sachen?_____

_____

38. Wer hatte nach fünf Wochen genug Material für den Aufsatz?___

_____

APPENDIX C

COMPUTED SIGNIFICANT WORDS

Word	Number	Significance Coefficient	Meier Frequency	Text Frequency	Word Number
Kamel	1	983.39	16	9	16, 43, 106, 195, 317, 420, 490, 514, 561
England	2	388.50	9	2	84, 125
Zoo	3	388.50	9	2	172, 272
badete	4	194.25	9	1	87
Badewanne	5	194.25	9	1	68
beendet	6	194.25	9	1	158
Boston	7	194.25	9	1	281
Cricket	8	194.25	9	1	99
Freizeit	9	194.25	9	1	102
hereinkommen	10	194.25	9	1	399
Kamelen	11	194.25	9	1	170
Kamels	12	194.25	9	1	355
Manuskriptes	13	194.25	9	1	558
Oase	14	194.25	9	1	80
Paris	15	194.25	9	1	162

Word	Number	Significance Coefficient	Meier Frequency	Text Frequency	Word Number
Plauderei	16	194.25	9	1	189
Preisausschreiben	17	194.25	9	1	382
Public Library	18	194.25	9	1	300
Sahara	19	194.25	9	1	74
Schaukelstuhl	20	194.25	9	1	339
Teetasse	21	194.25	9	1	70
Tennis	22	194.25	9	1	97
truimphierender	23	194.25	9	1	549
Überfliegen	24	194.25	9	1	322
Zehenspitzen	25	194.25	9	1	430
gebeugtem	26	174.82	10	1	481
Notizbuch	27	124.87	14	1	331
Studierzimmer	28	124.87	14	1	392
skizziert	29	109.26	16	1	196
Kamele	30	102.83	17	1	145
rauchte	31	102.83	17	1	452

Word	Number	Significance Coefficient	Meier Frequency	Text Frequency	Word Number
reserviert	32	102.83	17	1	261
Rauchwolke	33	97.12	18	1	473
arrangiert	34	92.01	19	1	341
nahes	35	92.01	19	1	202
verbracht	36	87.41	20	1	173
Titelblatt	37	83.25	21	1	555
Cafe	38	76.01	23	1	203
Amerikaner	39	71.35	49	2	27, 222
gekrümmtem	40	62.43	28	1	478
auszuarbeiten	41	58.27	30	1	208
Unbewussten	42	56.39	31	1	565
Information	43	54.63	32	1	305
amüsiert	44	46.00	38	1	183
Klubs	45	46.00	38	1	244
dramatischer	46	44.82	39	1	214
Schlafrock	47	42.64	41	1	405

Word	Number	Significance Coefficient	Meier Frequency	Text Frequency	Word Number
traute	48	39.73	44	1	58
Aufsatz	49	36.42	192	4	40, 119, 230, 296
Situationen	50	35.67	49	1	215

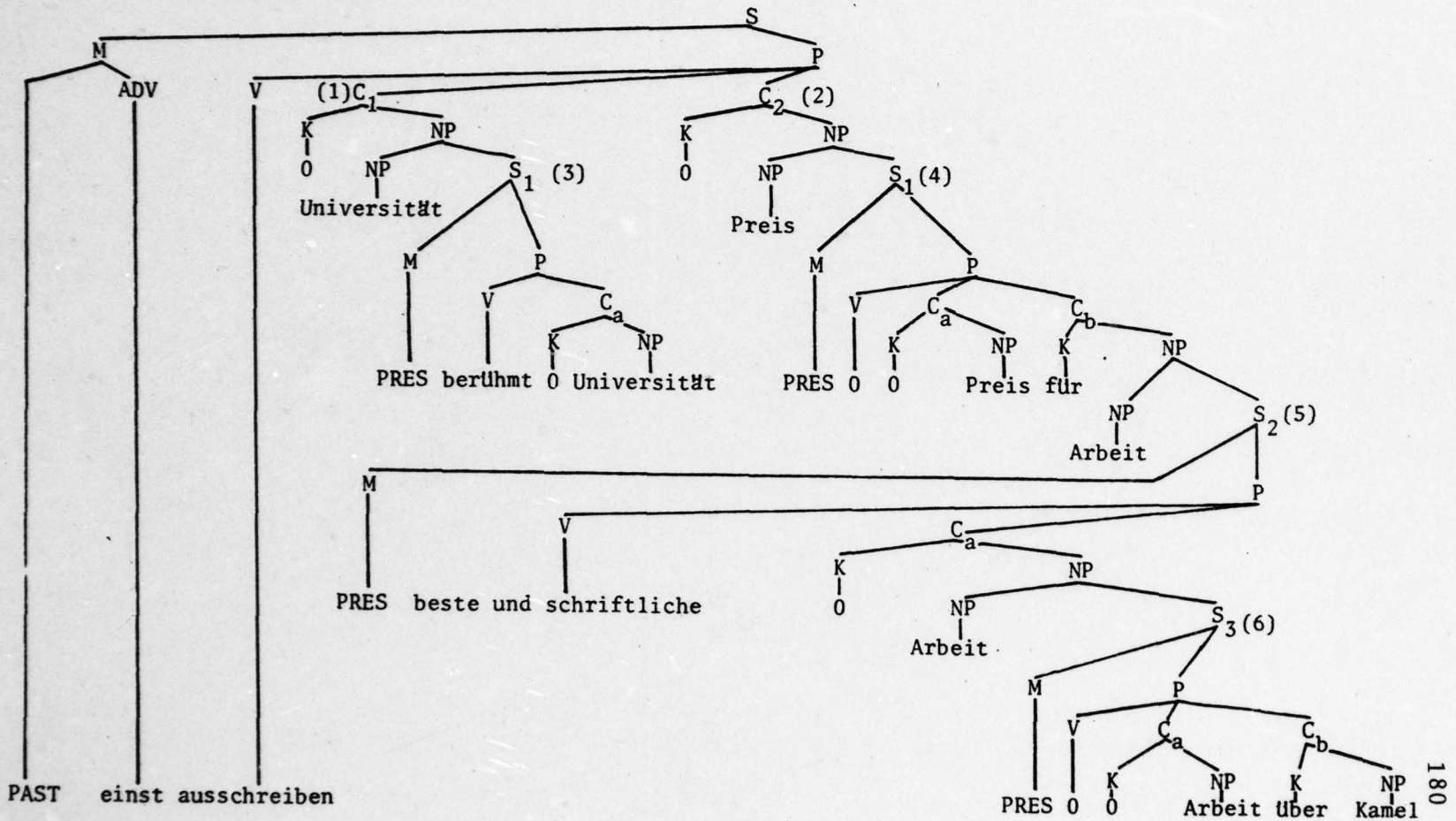
APPENDIX D

SENTENCE ANALYSES

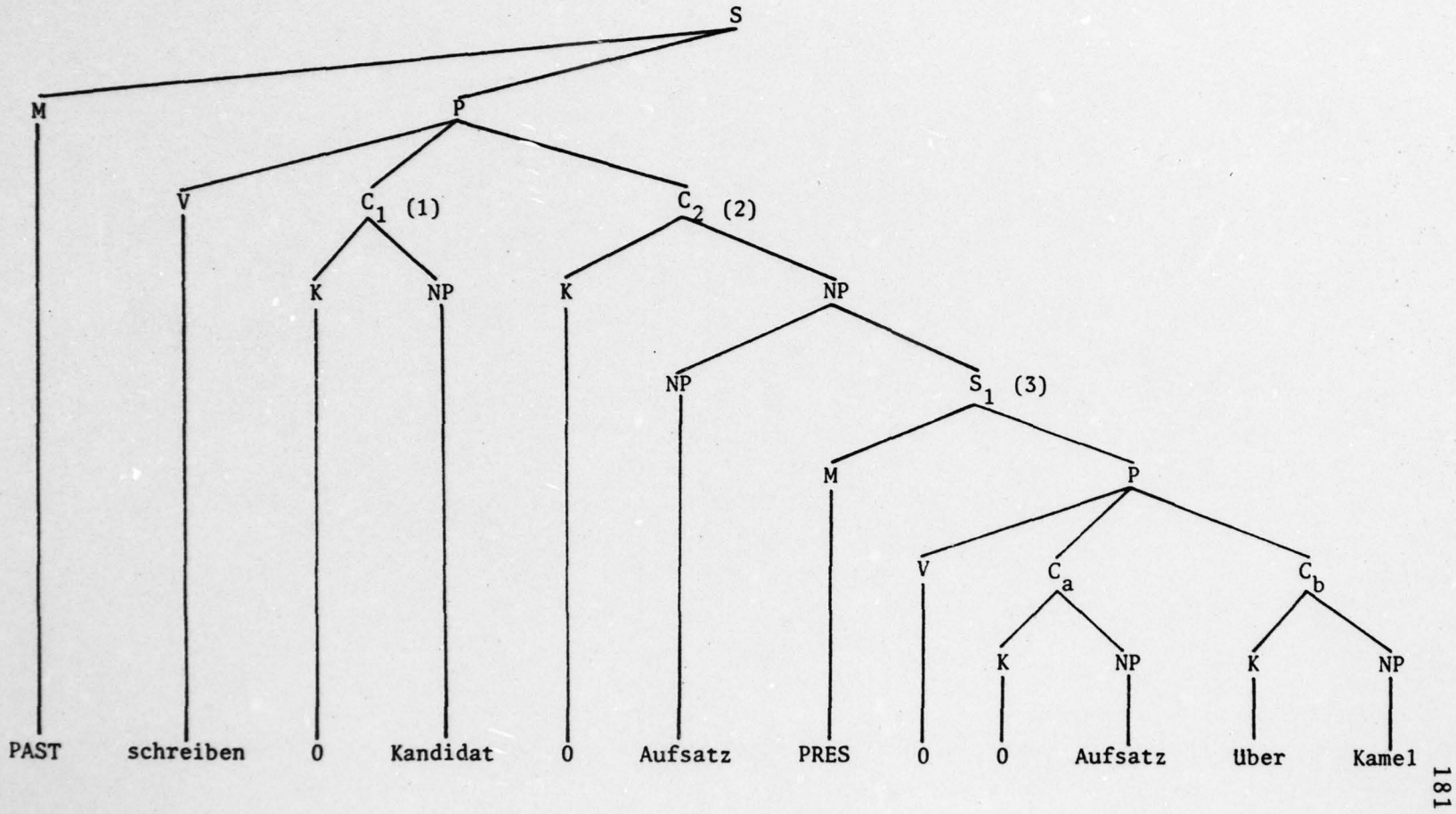
Explanation

Included along with each sentence analysis in Appendix D are the following four items: 1) the prepared (modified) sentence (S) from which the analysis was made, with the computed significant word underlined; 2) the question (Q) resulting from the sentence analysis; 3) the formally correct answer (A) to the question (this answer is based on the generic response); and 4) four numbers found under A through D in which A indicates the significance ranking of the underlined significant word, B the occurrence order of each significant word, C the word number of the significant word within the text, and D the constituent number that was randomly selected to generate the question.

S: Einst schrieb eine berühmte Universität einen Preis aus für die beste schriftliche Arbeit über das Kamel.
Q: Was schrieb einst eine berühmte Universität aus?
A: Einen Preis für die beste schriftliche Arbeit über das Kamel.
A = 1 B = 1 C = 16 D = (2).

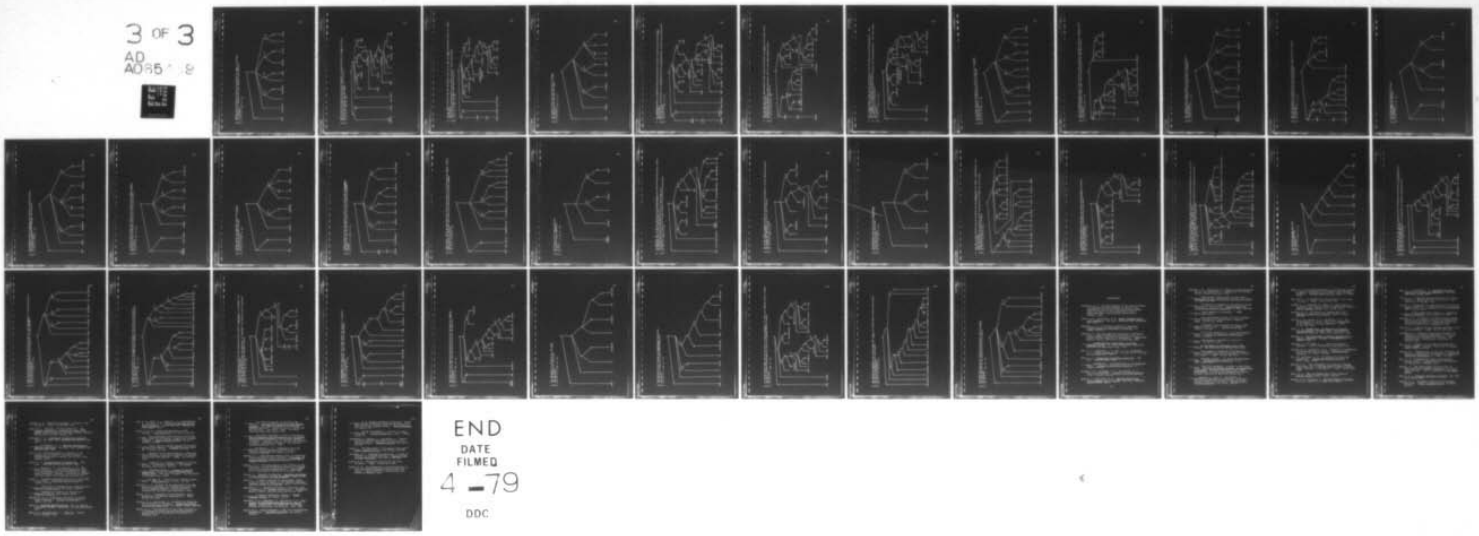


S: Jeder Kandidat schrieb einen Aufsatz über das Kamel.
Q: Wer schrieb einen Aufsatz über das Kamel?
A: Jeder Kandidat.
A = 1 B = 2 C = 43 D = (1).

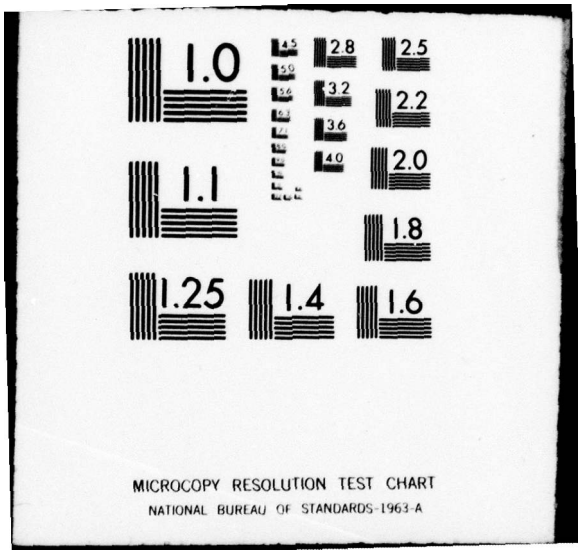


AD-A065 489 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO F/G 5/10
THE CLOZE TEST AS A PROCEDURE FOR ESTABLISHING OBJECTIVE GERMAN--ETC(U)
1978 B M STRONG
UNCLASSIFIED AFIT-CI-79-100D NL

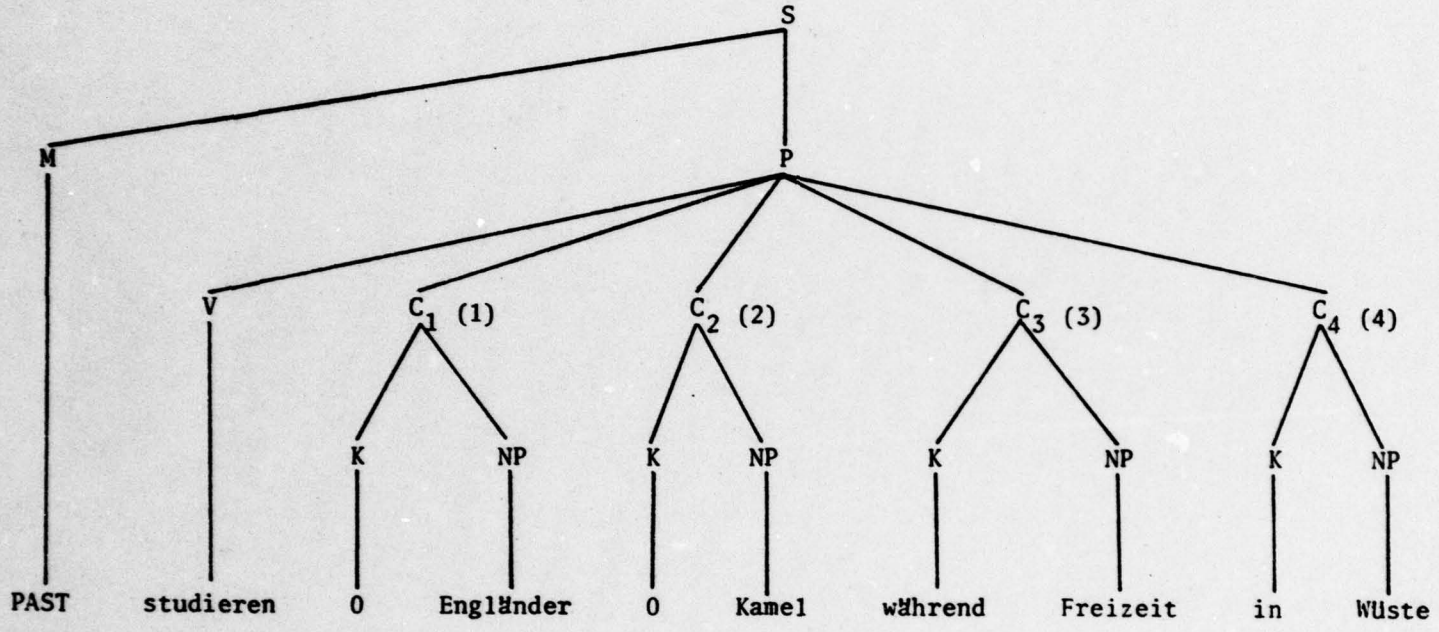
3 OF 3
AD
A065 489



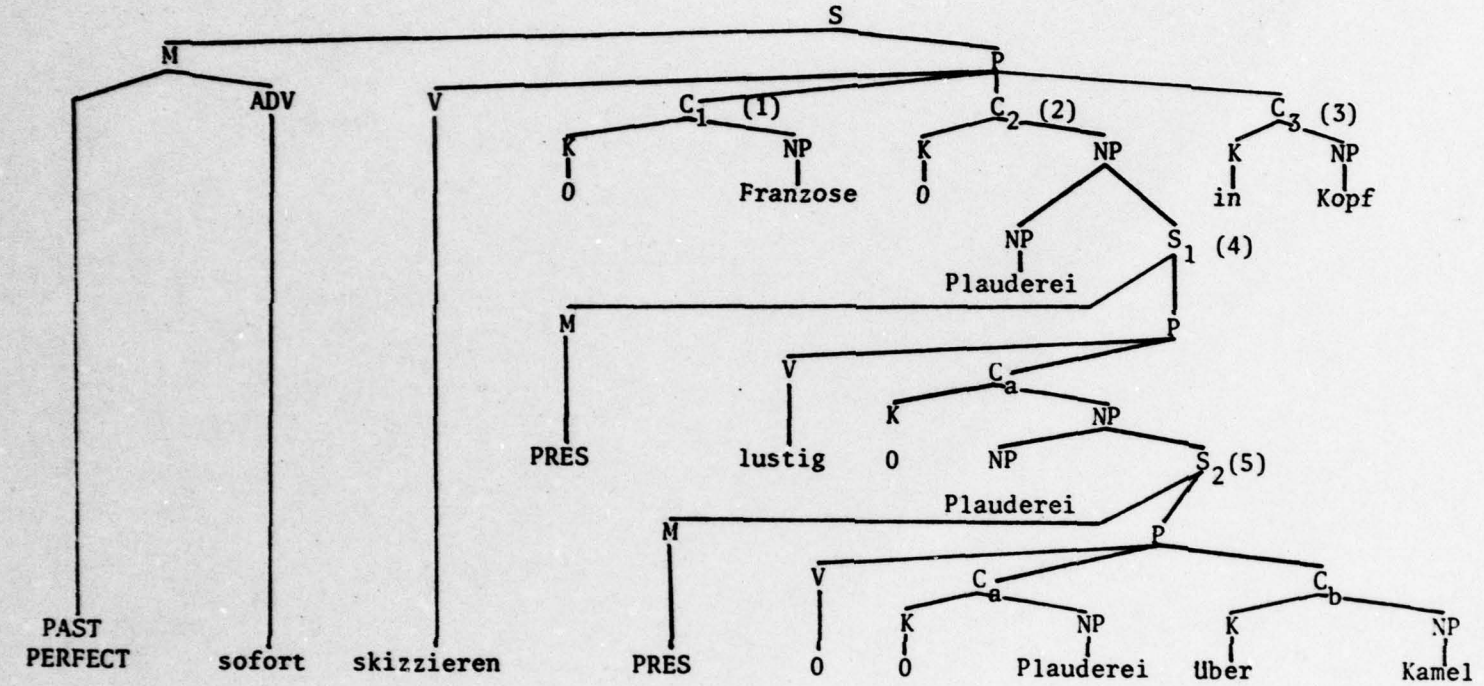
END
DATE
FILMED
4-79
DDC


1.0
1.1
1.25
1.4
1.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.8
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

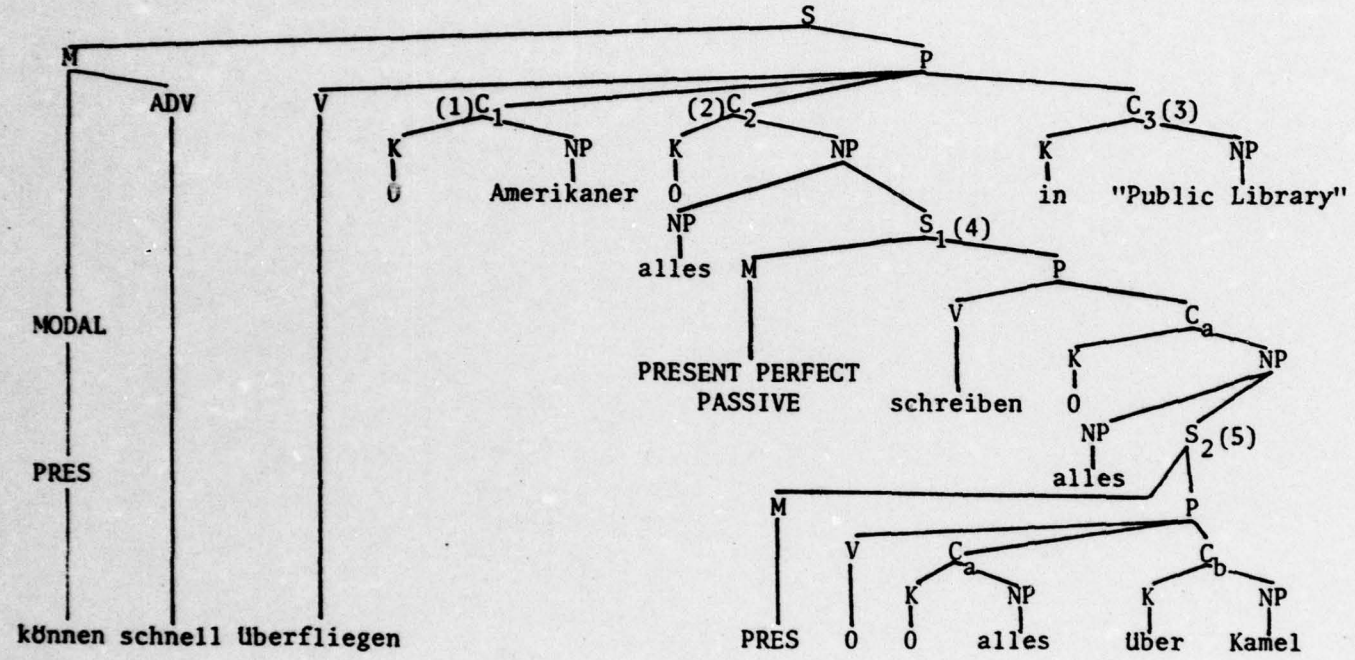
S: Während seiner Freizeit studierte der Engländer das Kamel in der Wüste.
Q: Wo studierte der Engländer das Kamel während seiner Freizeit?
A: In der Wüste.
A = 1 B = 3 C = 106 D = (4).



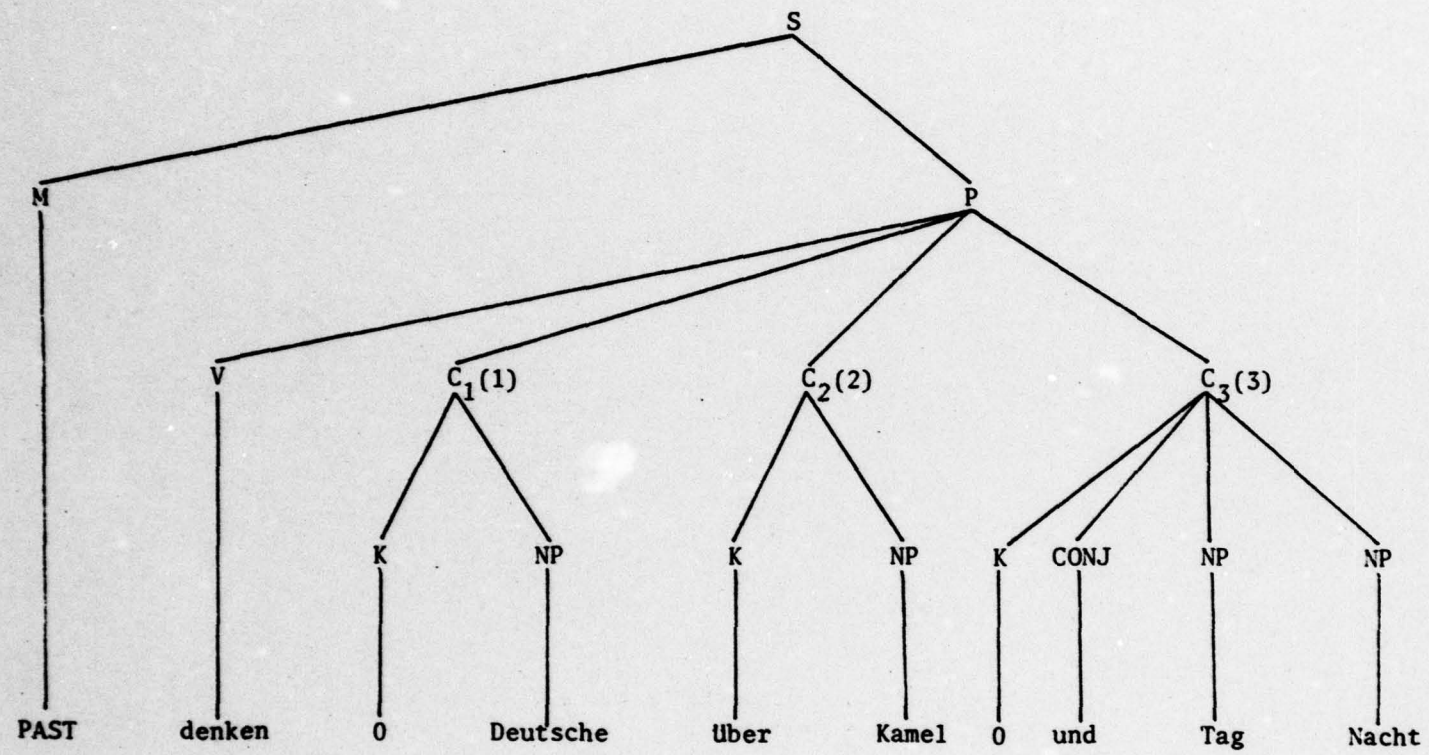
S: Sofort hatte der Franzose eine lustige Plauderei in seinem Kopf über das Kamel skizziert.
Q: Was hatte der Franzose in seinem Kopf sofort skizziert?
A: Eine lustige Plauderei über das Kamel.
A = 1 B = 4 C = 195 D = (2).



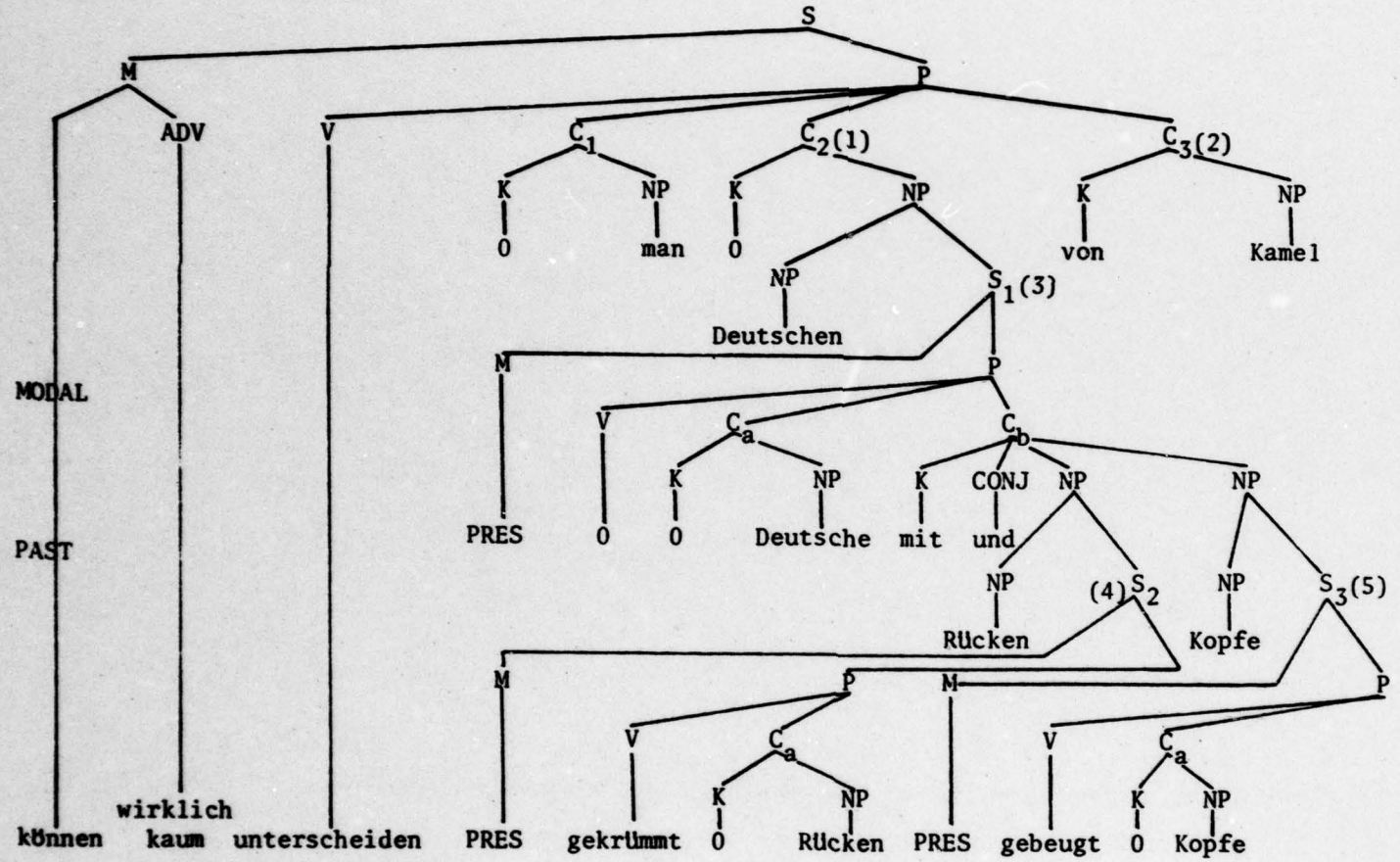
S: In der "Public Library" kann der Amerikaner alles, was je über das Kamel geschrieben worden ist, schnell überfliegen.
Q: Was alles kann der Amerikaner in der "Public Library" schnell überfliegen?
A: Alles, was je über das Kamel geschrieben worden ist.
A = 1 B = 5 C = 317 D = (4).



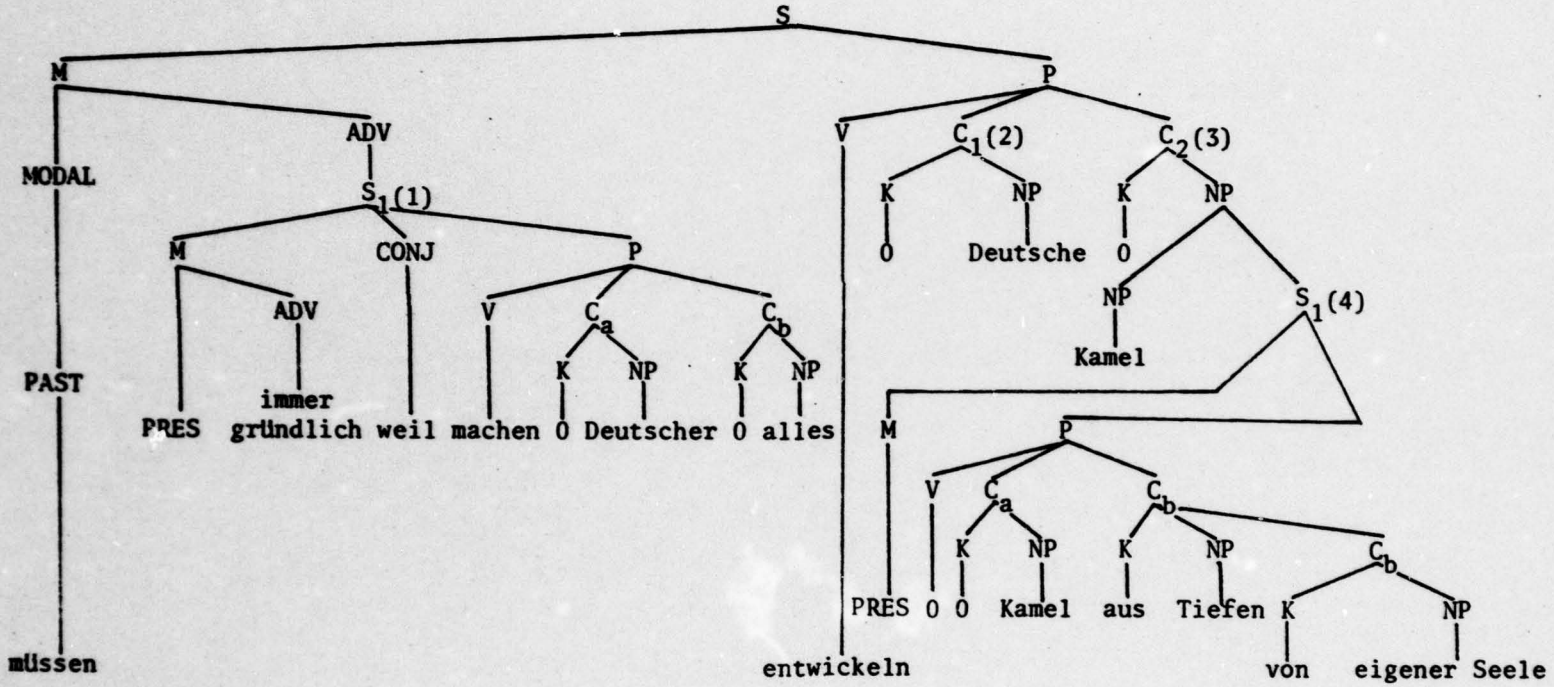
S: Der Deutsche dachte Tag und Nacht über das Kamel.
Q: Wer dachte Tag und Nacht über das Kamel?
A: Der Deutsche.
A = 1 B = 6 C = 420 D = (1).



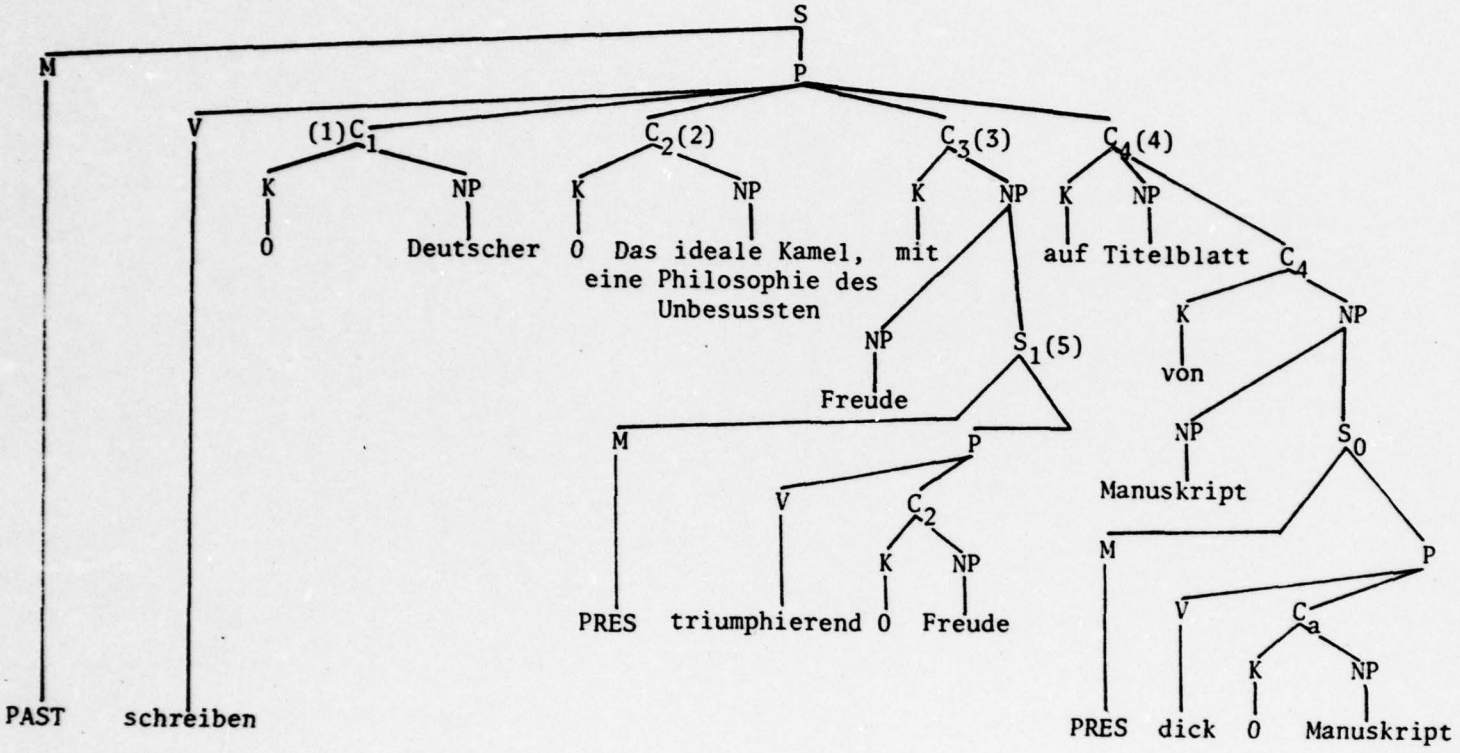
S: Mit gekrümmtem Rücken und gebeugtem Kopfe konnte man den Deutschen wirklich kaum von einem Kamel unterscheiden.
Q: Wovon konnte man den Deutschen, mit gekrümmtem Rücken und gebeugtem Kopfe, wirklich kaum unterscheiden?
A: Von einem Kamel.
A = 1 B = 7 C = 490 D = (2).



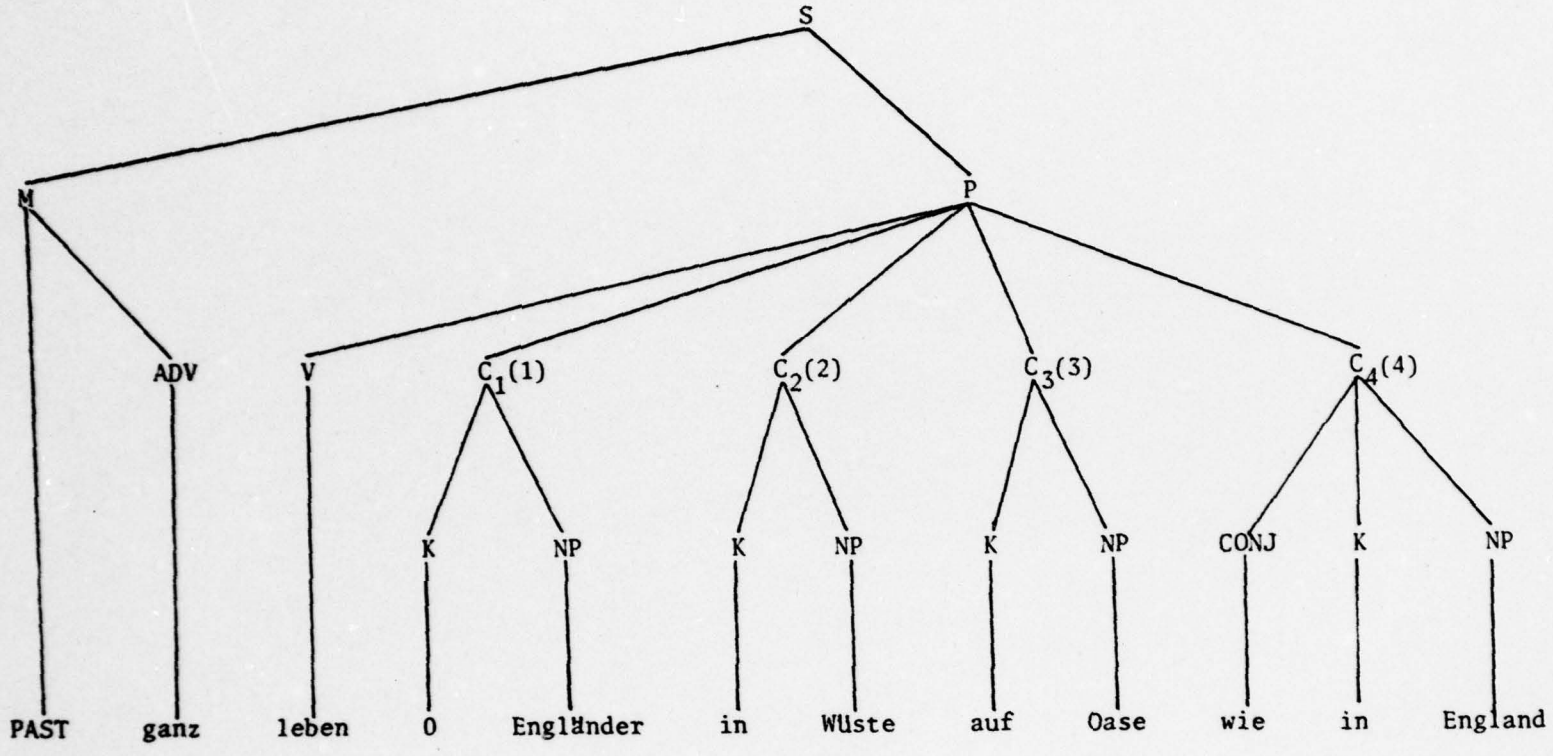
S: Weil ein Deutscher alles immer gründlich macht, musste der Deutsche das Kamel aus den Tiefen seiner eigenen Seele entwickeln.
Q: Wie musste der Deutsche das Kamel entwickeln, weil er alles immer gründlich macht?
A: Aus den Tiefen seiner eigenen Seele.
A = 1 B = 8 C = 514 D = (4).



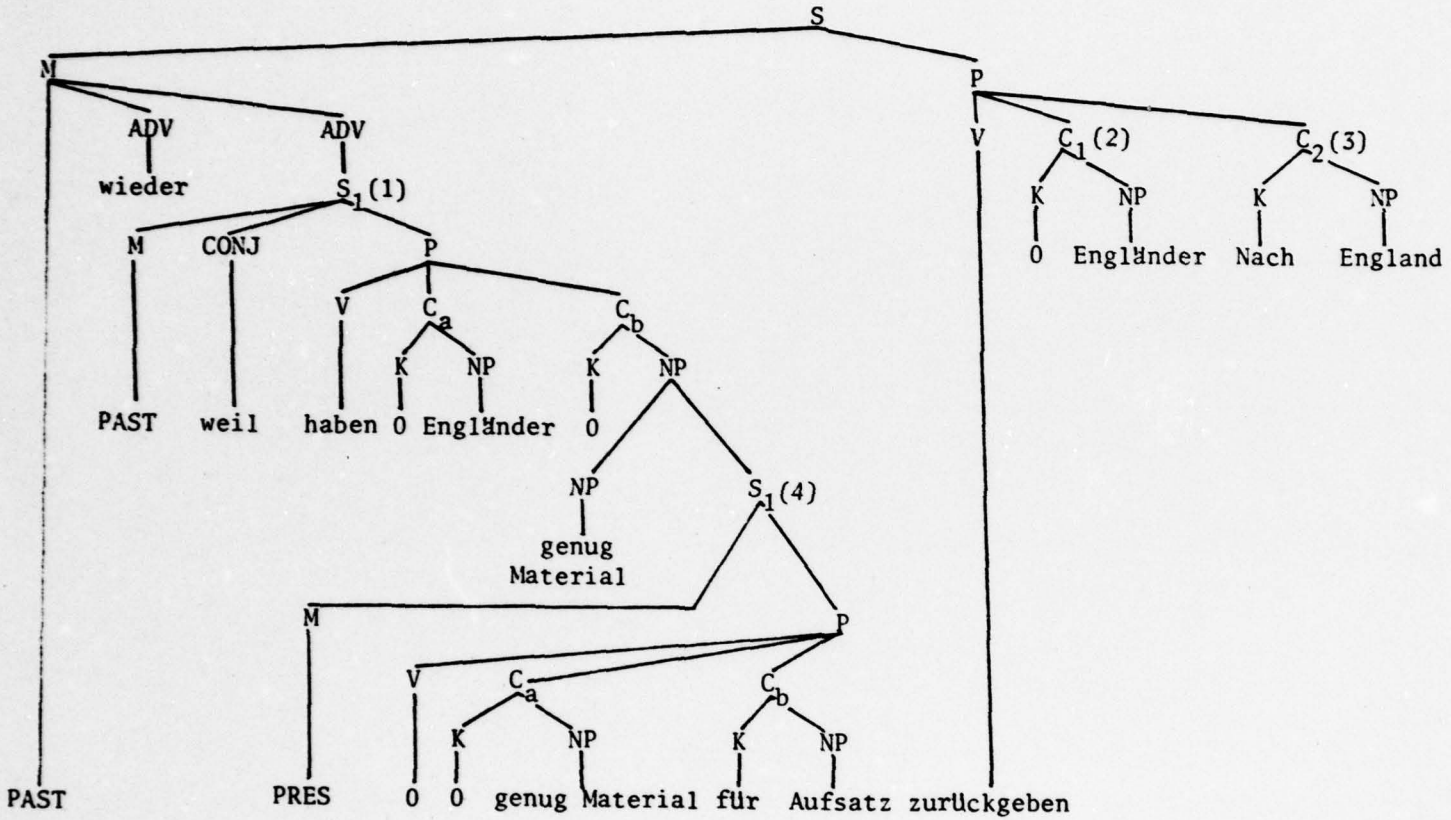
S: Mit triumphierender Freude schrieb der Deutsche auf das Titelblatt des dicken Manuskriptes: "Das ideale Kamel, eine Philosophie des Unbewussten."
Q: Wer schrieb mit triumphierender Freude auf das Titelblatt des dicken Manuskriptes: "Das ideale Kamel, eine Philosophie des Unbewussten?"
A: Der Deutsche.
A = 1 B = 9 C = 561 D = (1).



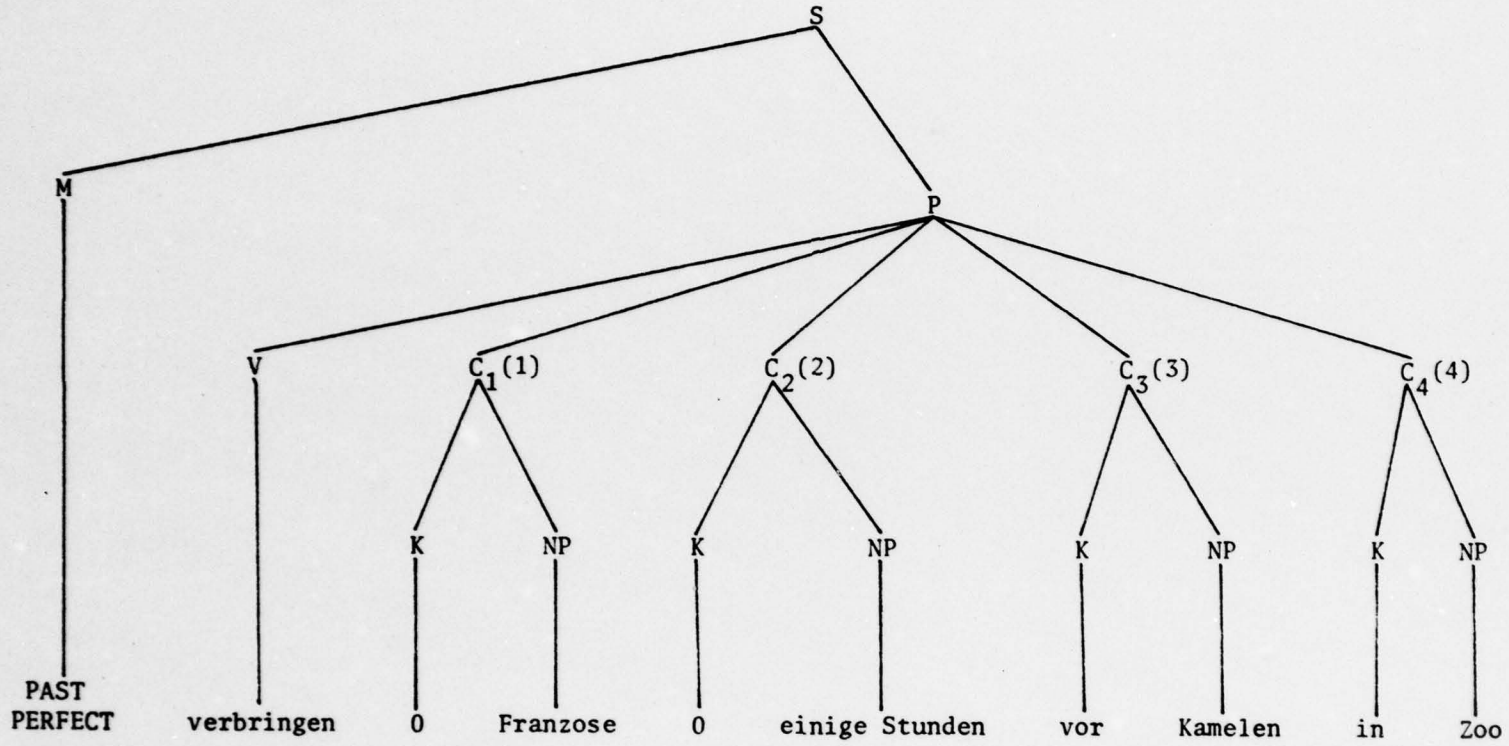
S: In der Wüste lebte der Engländer auf einer Oase ganz wie in England.
Q: Wo lebte der Engländer in der Wüste ganz wie in England?
A: Auf einer Oase.
A = 2 B = 1 C = 84 D = (3).



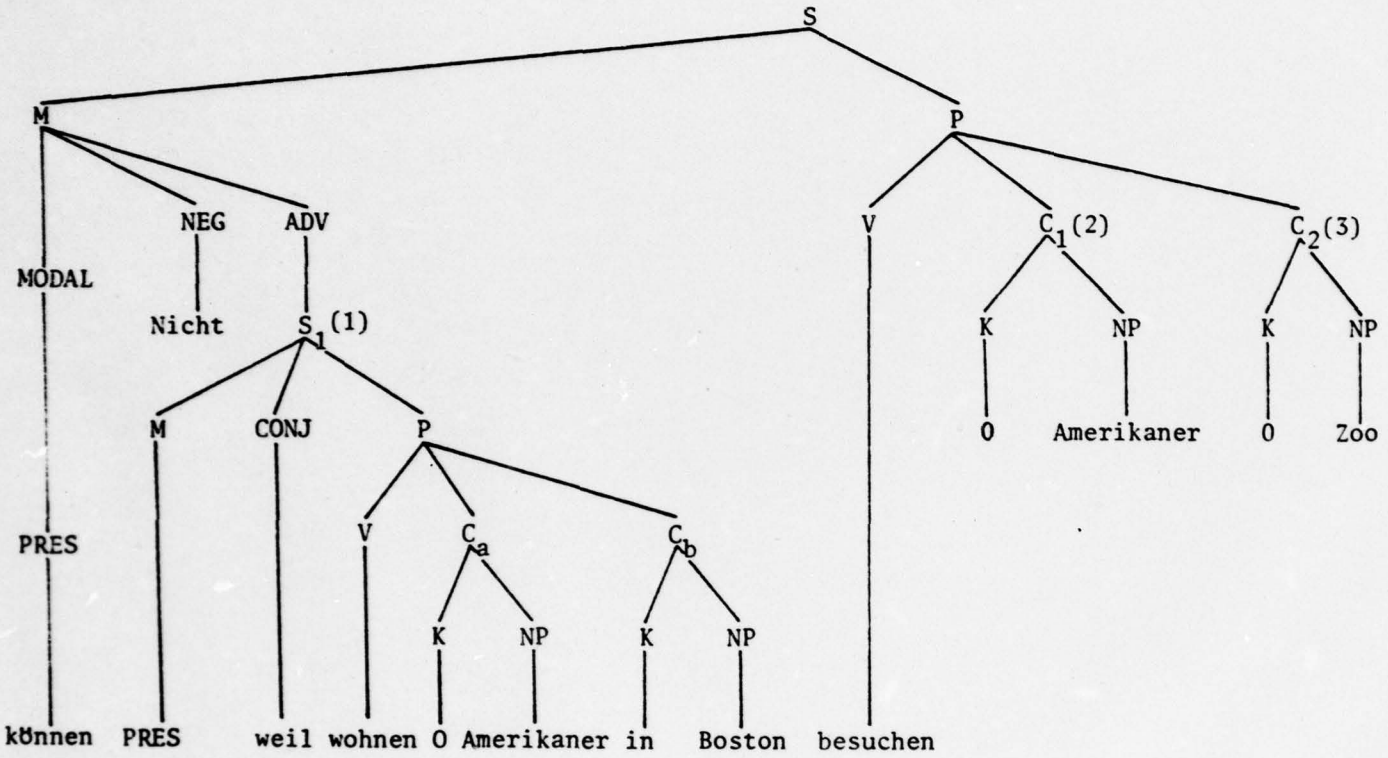
S: Weil der Engländer genug Material für den Aufsatz hatte, ging der Engländer wieder nach England zurück.
Q: Wohin ging der Engländer wieder zurück, weil er genug Material für den Aufsatz hatte?
A: Nach England.
A = 2 B = 2 C = 125 D = (3).



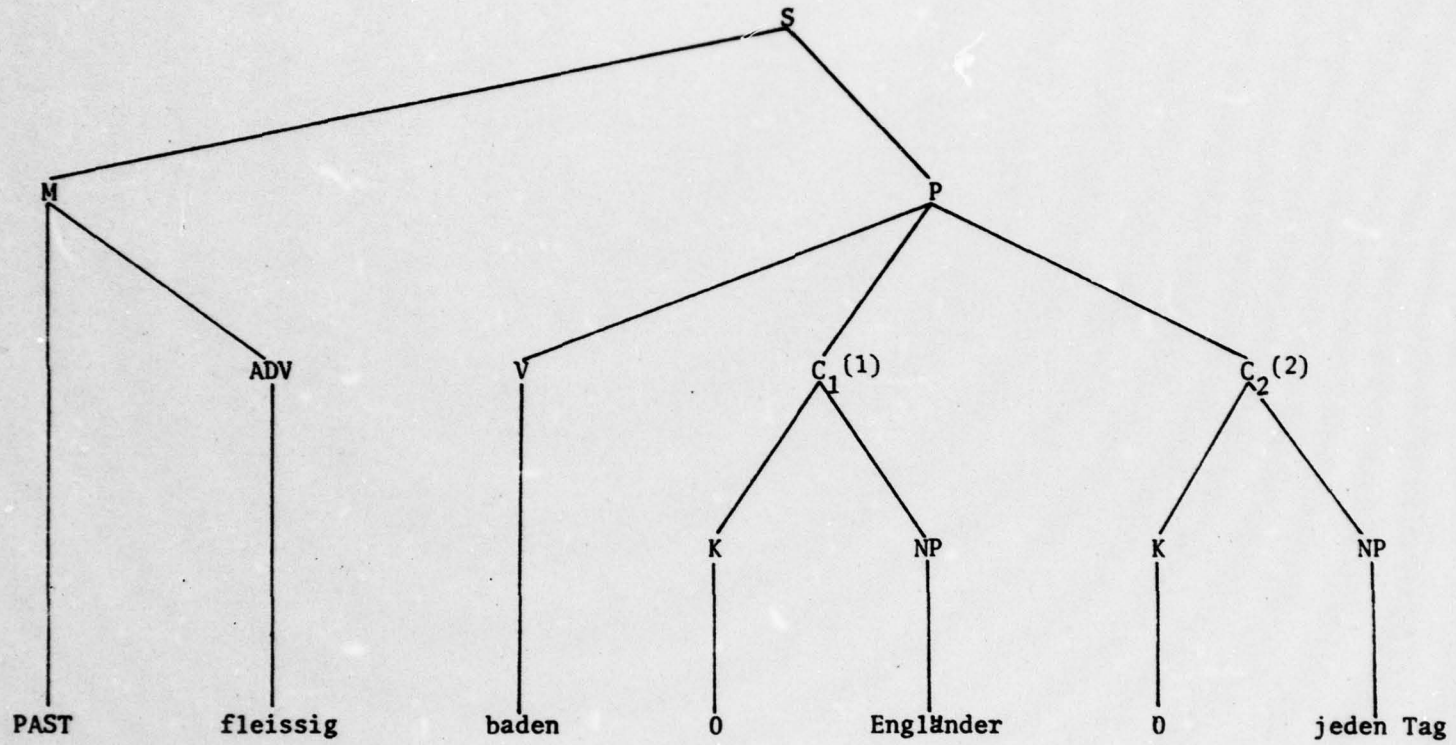
S: Der Franzose hatte einige Stunden vor den Kamelen im Zoo verbracht.
Q: Wie lange hatte der Franzose vor den Kamelen im Zoo verbracht?
A: Einige Stunden.
A = 3 B = 1 C = 172 D = (2).



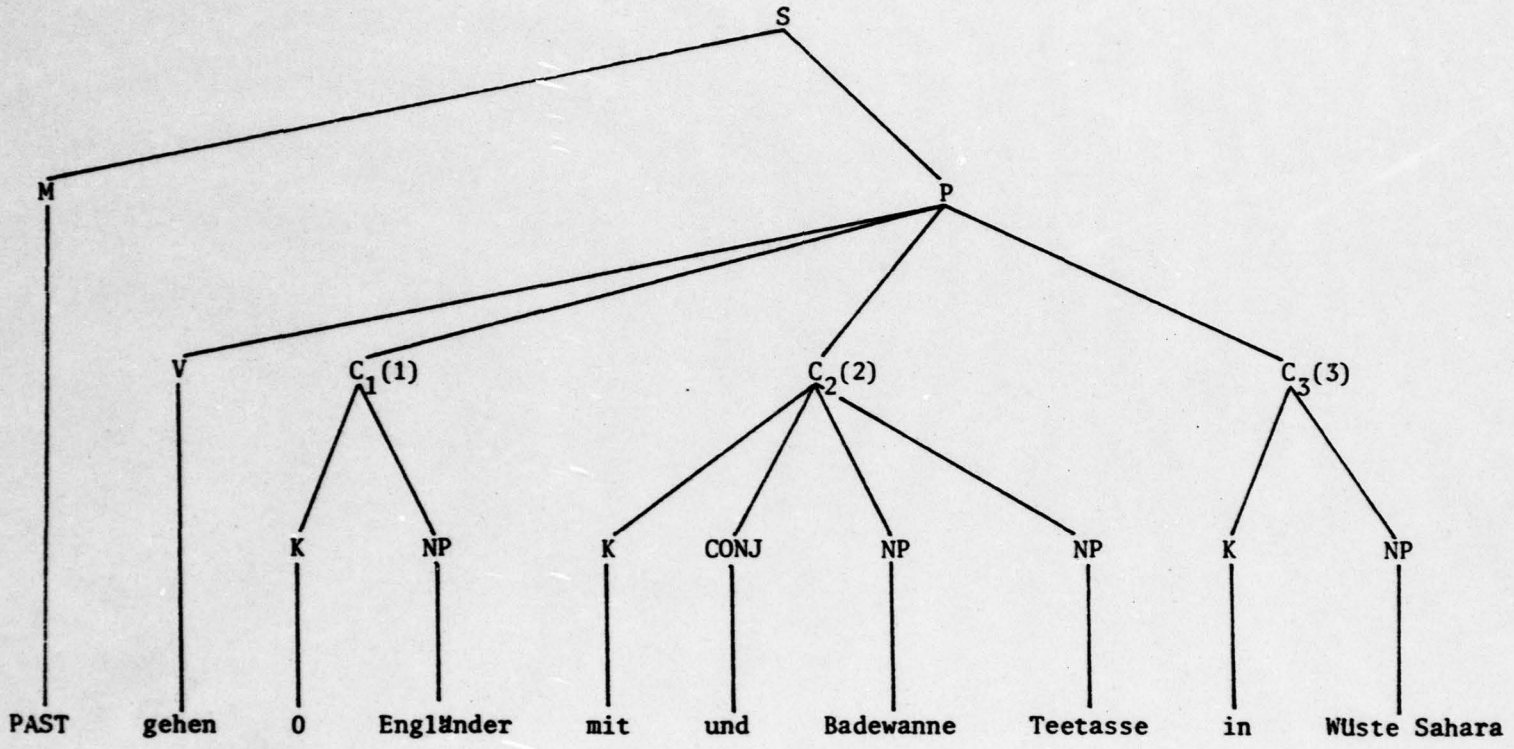
S: Weil der Amerikaner in Boston wohnt, kann der Amerikaner den Zoo auch nicht besuchen.
Q: Wer kann den Zoo nicht besuchen, weil er in Boston wohnt?
A: Der Amerikaner.
A = 3 B = 2 C = 272 D = (2).



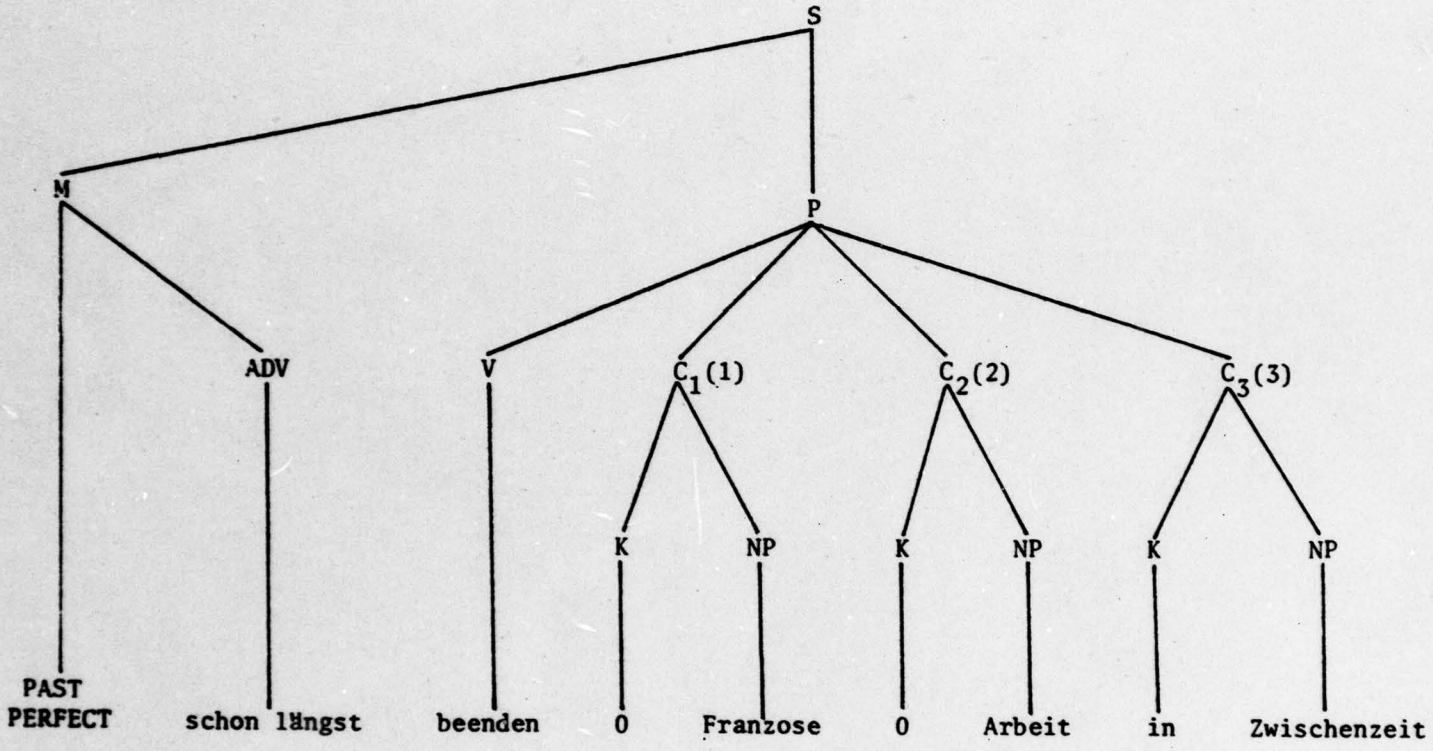
S: Jeden Tag badete der Engländer fleissig.
Q: Wer badete jeden Tag fleissig?
A: Der Engländer.
A = 4 B = 1 C = 87 D = (1).



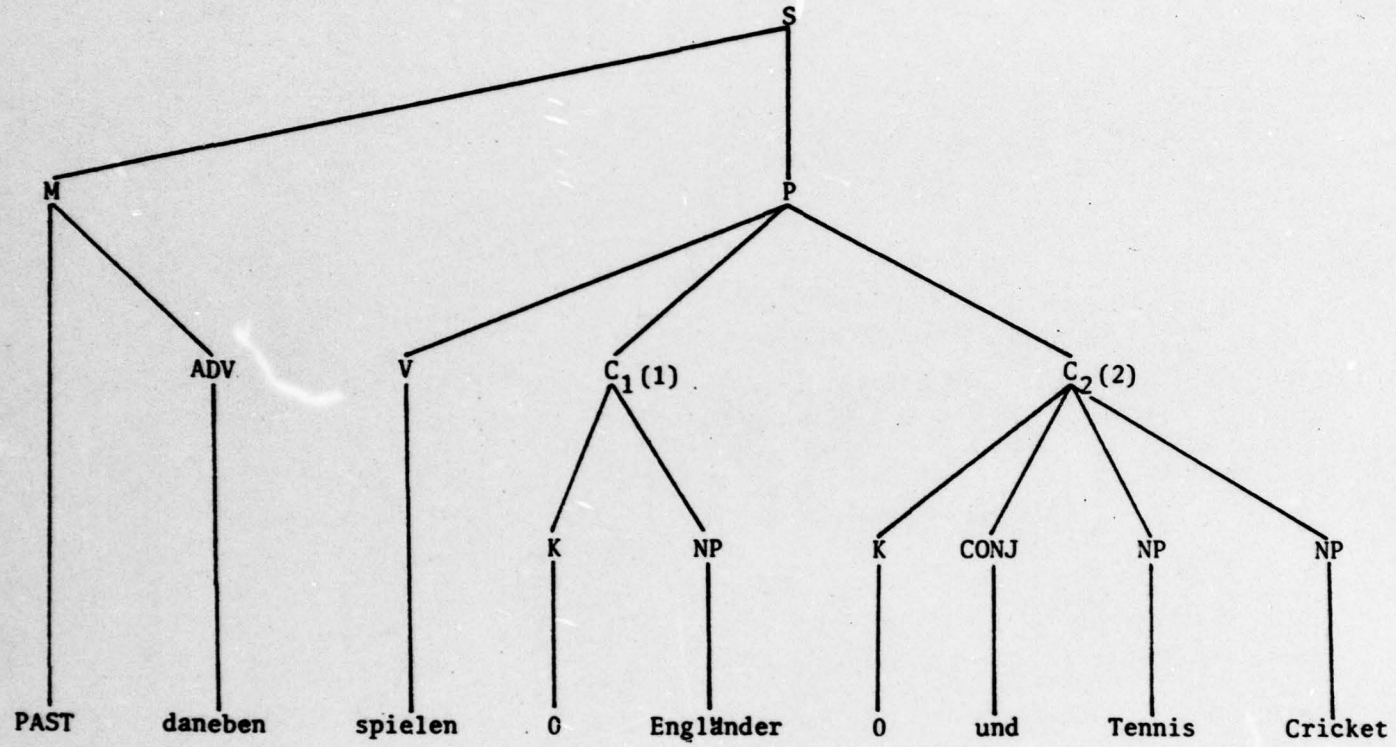
S: Der Engländer ging mit Badewanne und Teetasse in die Wüste Sahara.
Q: Wohin ging der Engländer mit Badewanne und Teetasse?
A: In die Wüste Sahara.
A = 5 B = 1 C = 68 D = (3).



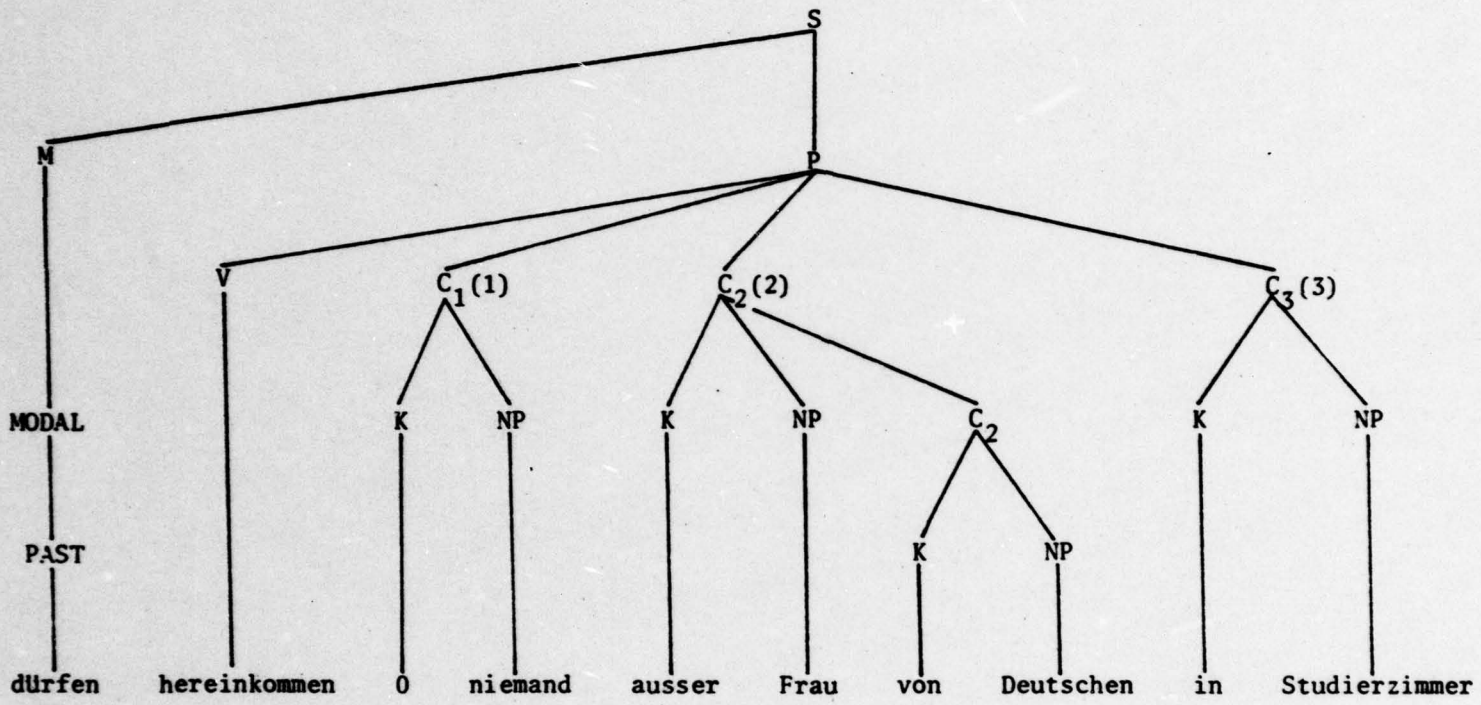
S: In der Zwischenzeit hatte der Franzose seine Arbeit schon längst beendet.
Q: Wer hatte in der Zwischenzeit seine Arbeit schon längst beendet?
A: Der Franzose.
A = 6 B = 1 C = 158 D = (1).



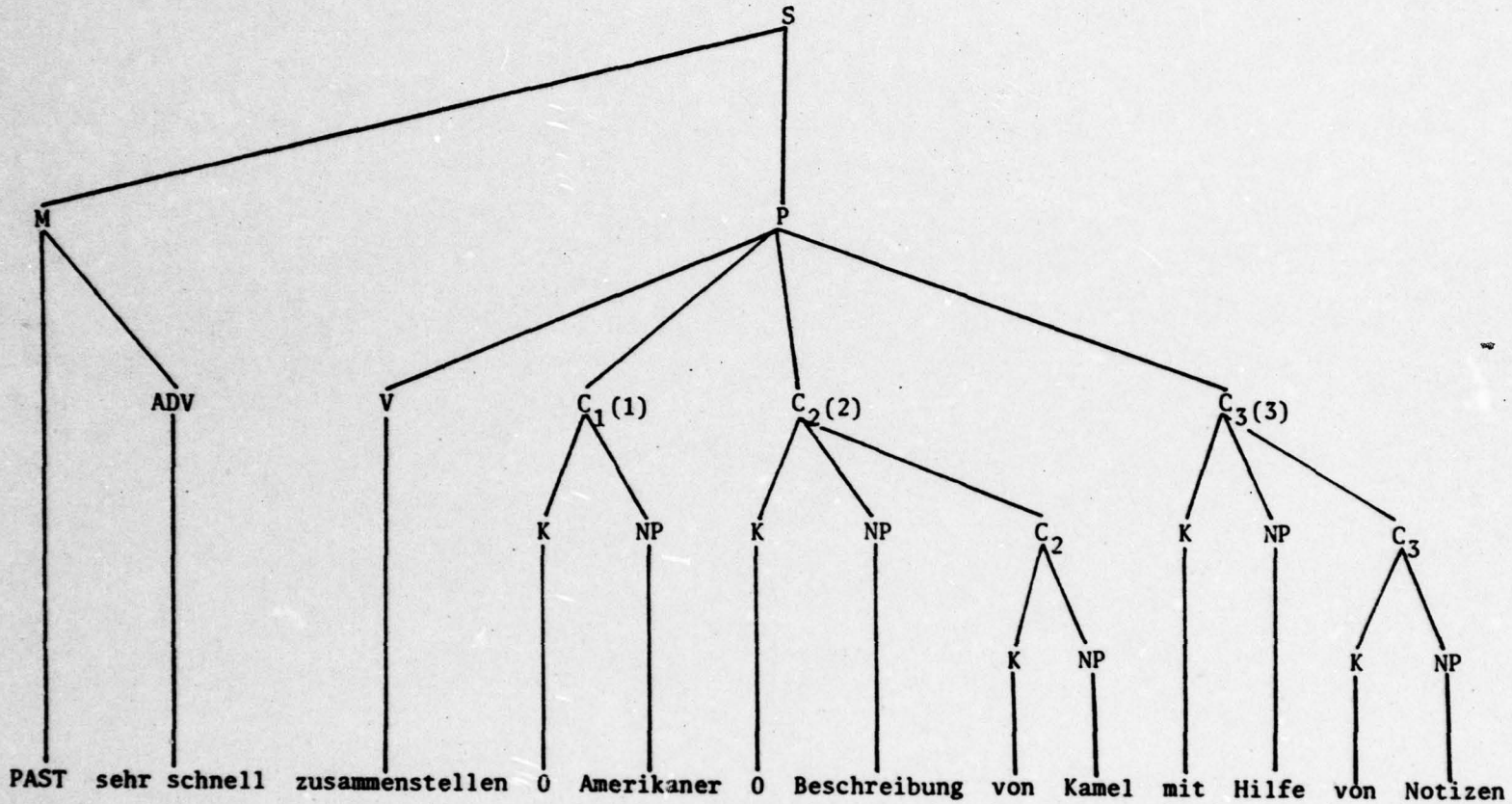
S: Daneben spielte der Engländer Tennis und Cricket.
Q: Was spielte der Engländer daneben?
A: Tennis und Cricket.
A = 8 B = 1 C = 99 D = (2).



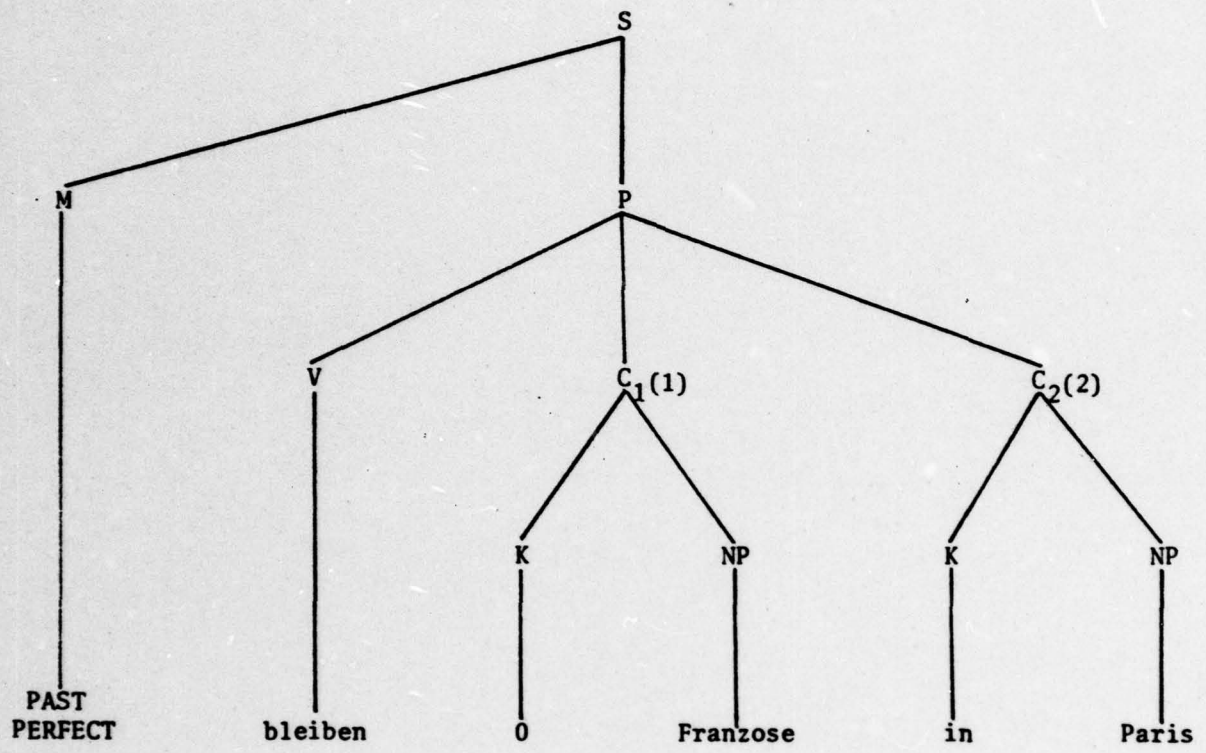
S: Niemand ausser der Frau des Deutschen durfte ins Studierzimmer hereinkommen.
Q: Wer ausser der Frau des Deutschen durfte ins Studierzimmer hereinkommen?
A: Niemand.
A = 10 B = 1 C = 399 D = (1).



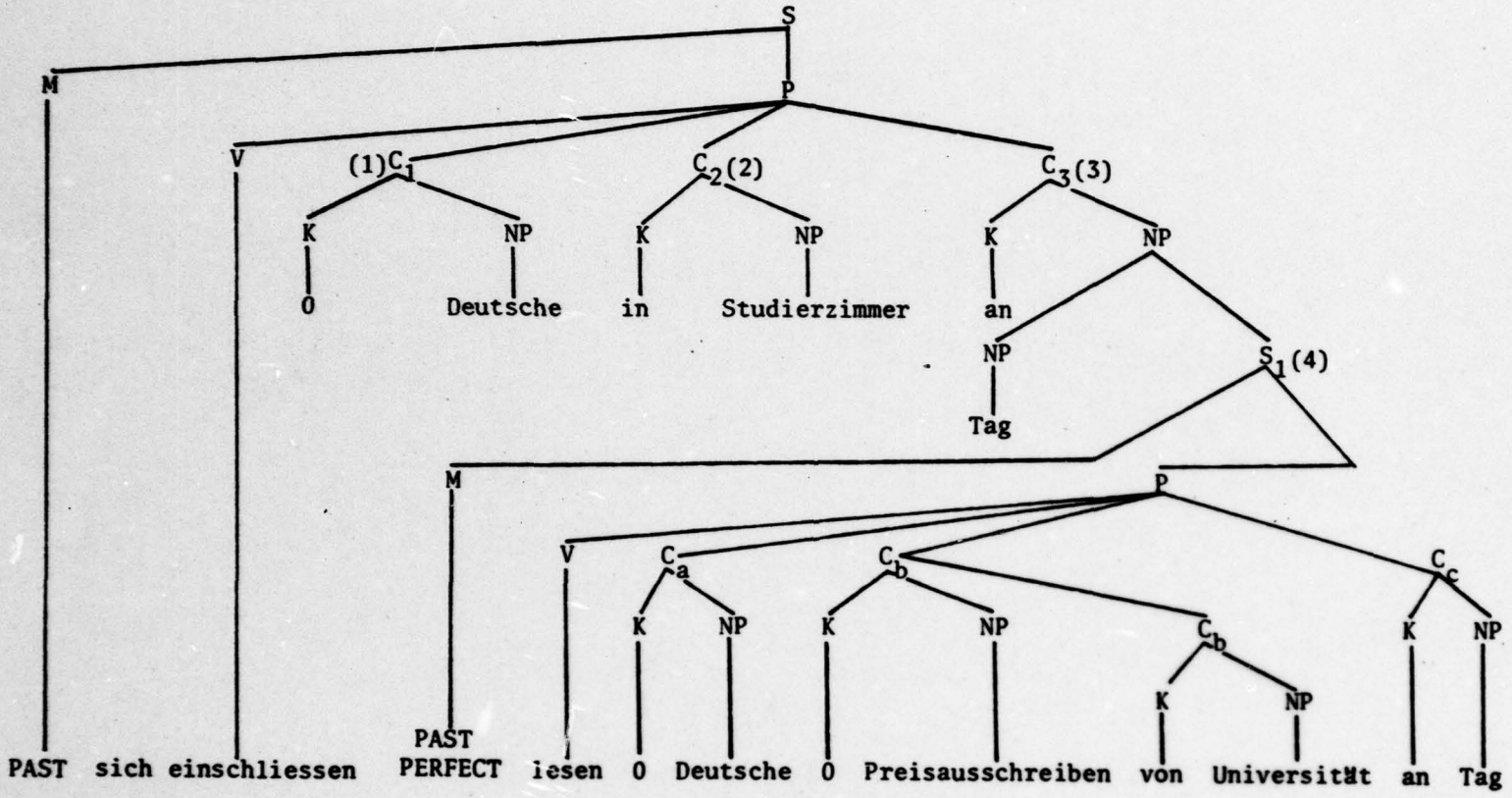
S: Sehr schnell stellte der Amerikaner mit Hilfe seiner Notizen eine Beschreibung des Kamels zusammen.
Q: Womit stellte der Amerikaner sehr schnell eine Beschreibung des Kamels zusammen?
A: Mit Hilfe seiner Notizen.
A = 12 B = 1 C = 355 D = (3).



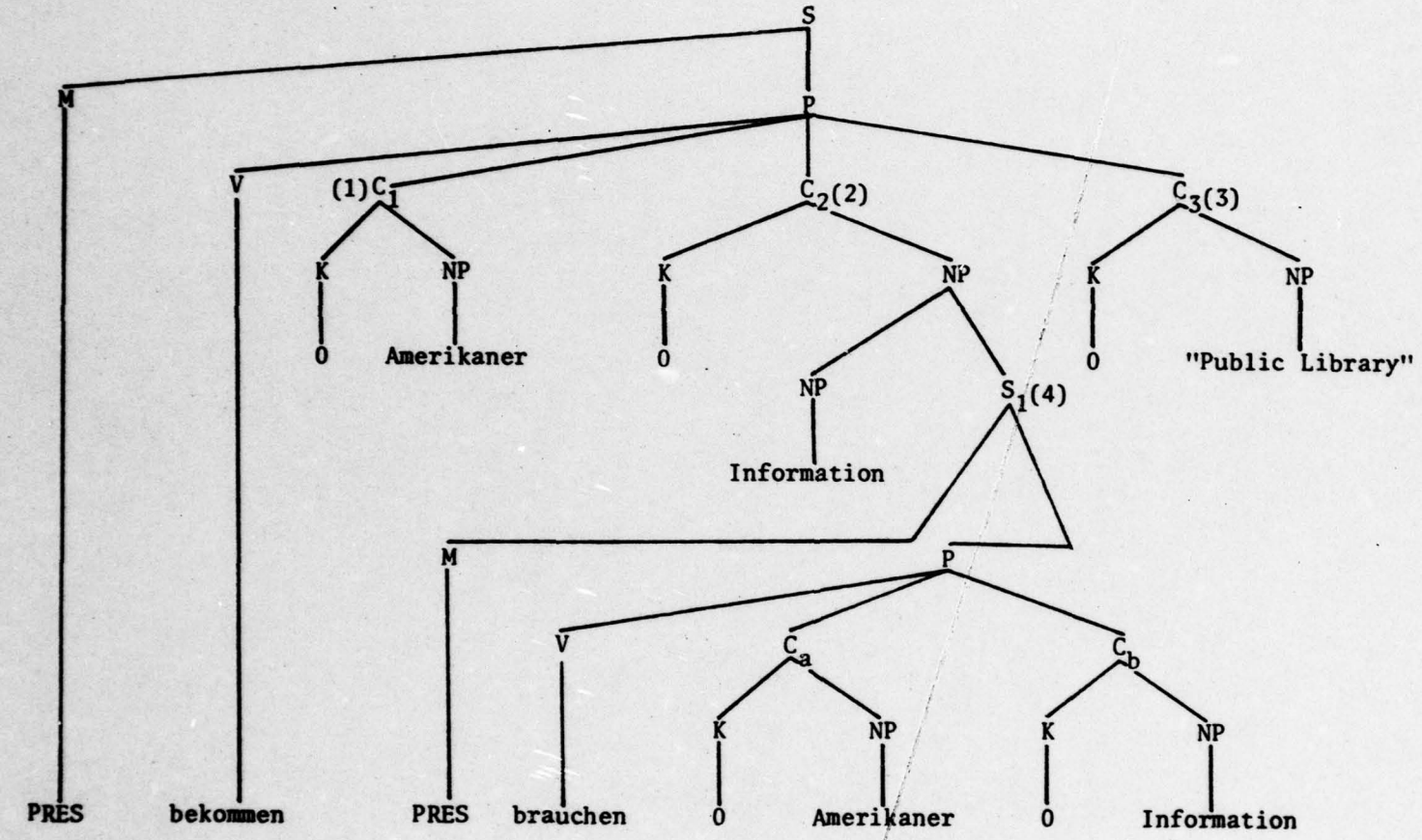
S: Der Franzose war in Paris geblieben.
Q: Wo war der Franzose geblieben?
A: In Paris.
A = 15 B = 1 C = 162 D = (2).



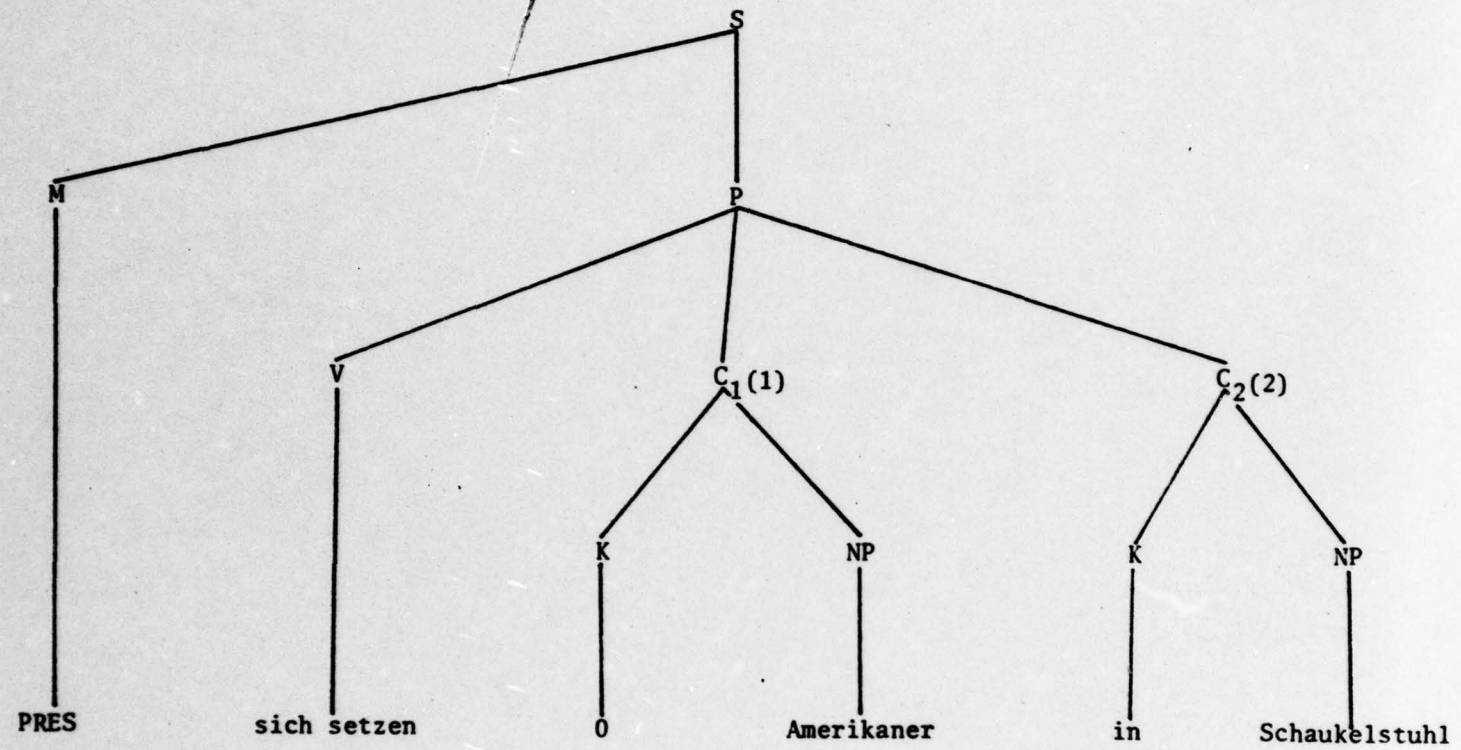
S: An demselben Tag, an dem der Deutsche das Preisausschreiben der Universität gelesen hatte, schloss der Deutsche sich in sein Studierzimmer ein.
Q: An welchem Tag schloss der Deutsche sich in sein Studierzimmer ein?
A: An demselben Tag, an dem er das Preisausschreiben der Universität gelesen hatte.
A = 17 B = 1 C = 382 D = (4).



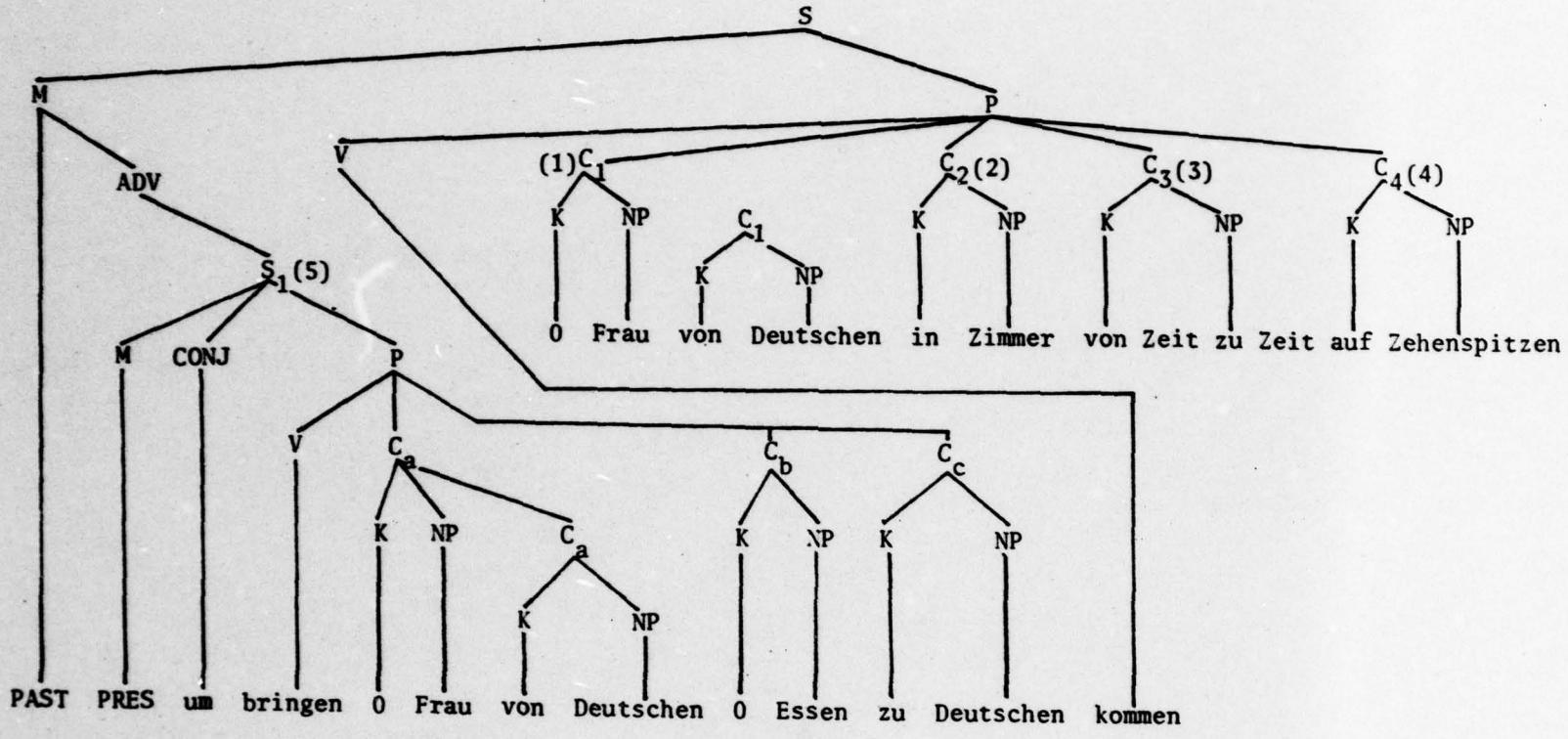
S: Nur in der "Public Library" bekommt der Amerikaner die Information, die der Amerikaner braucht.
Q: Was bekommt der Amerikaner nur in der "Public Library"?
A: Die Information, die er braucht.
A = 18 B = 1 C = 300 D = (2).



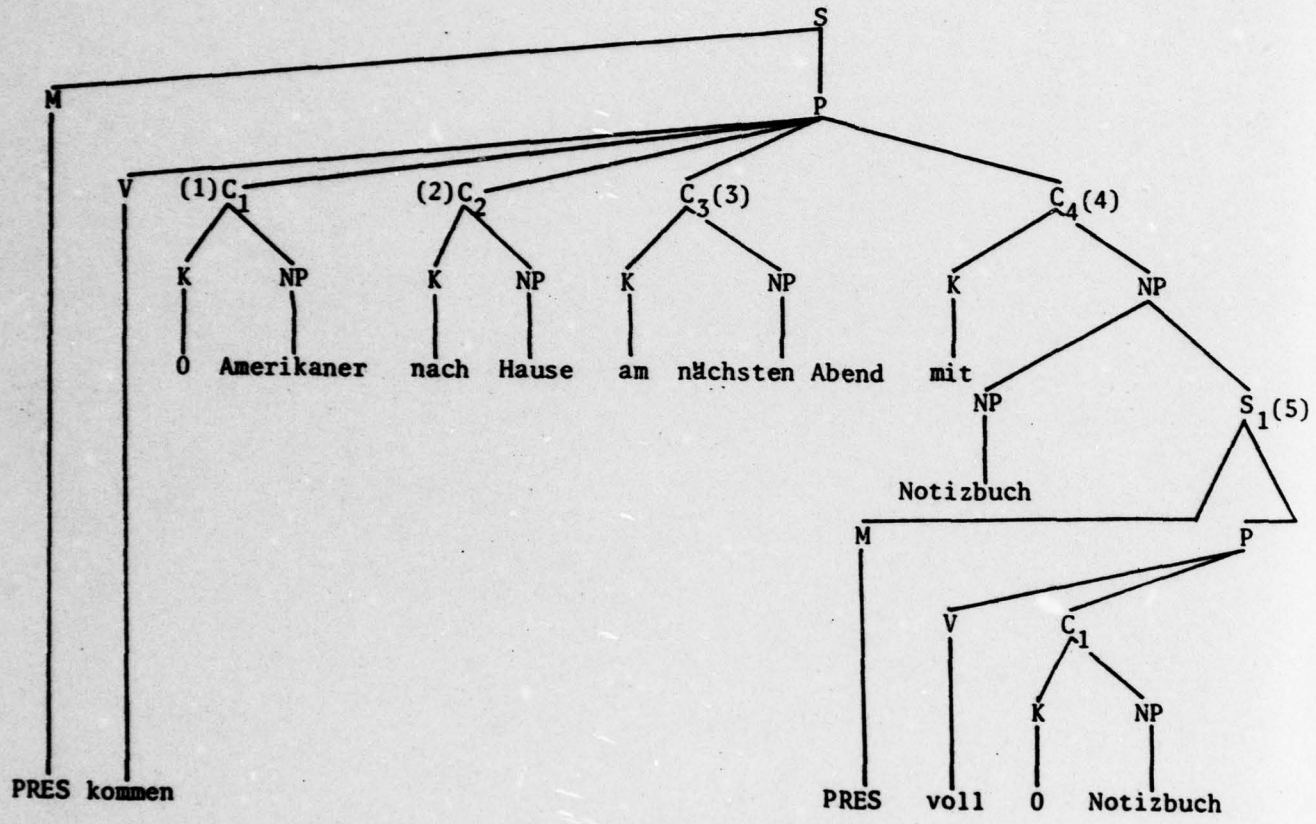
S: Der Amerikaner setzt sich in den Schaukelstuhl.
Q: Worin setzt sich der Amerikaner?
A: In den Schaukelstuhl.
A = 20 B = 1 C = 339 D = (2).



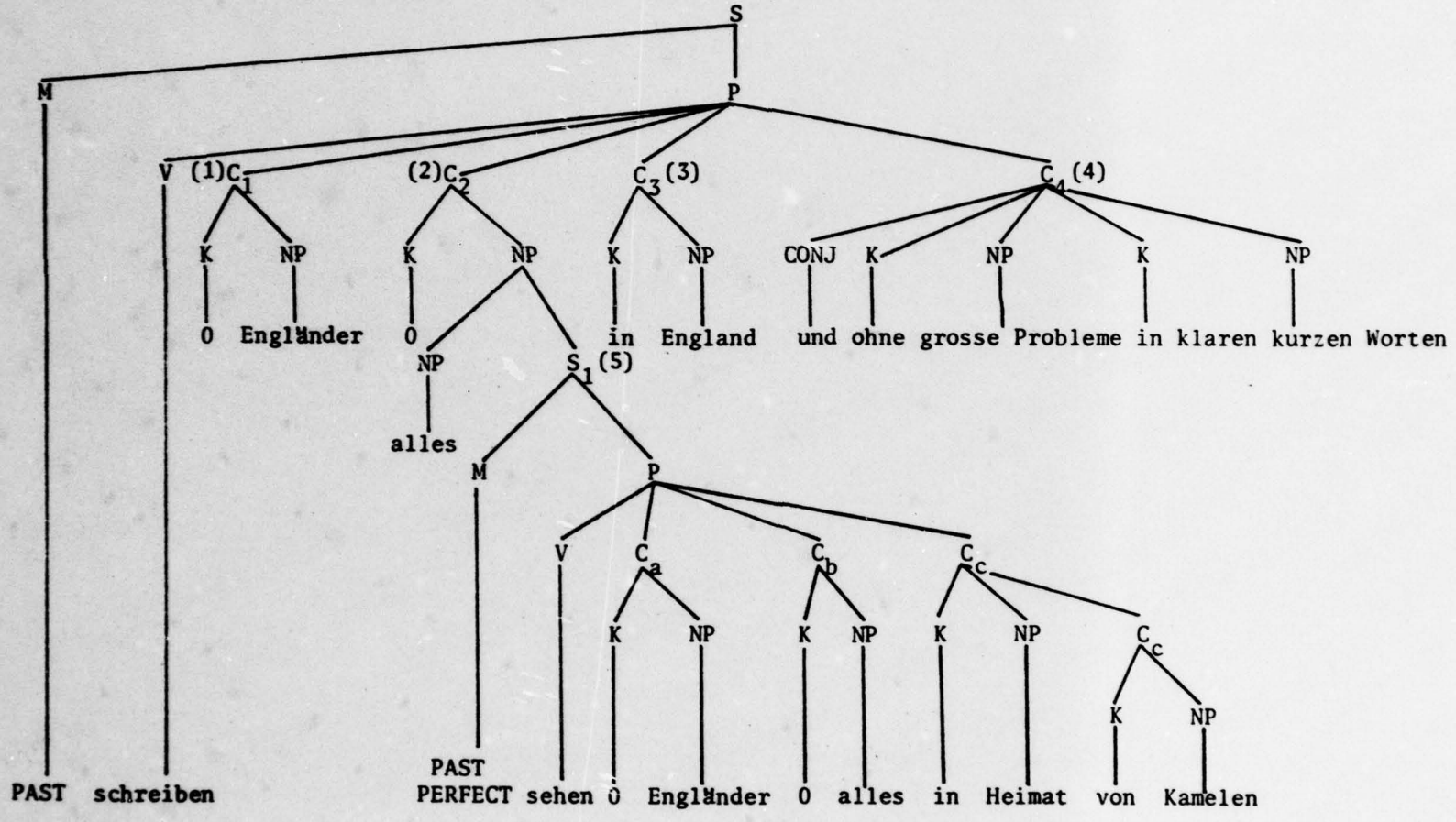
S: Von Zeit zu Zeit kam die Frau des Deutschen auf den Zehenspitzen ins Zimmer, um dem Deutschen sein Essen zu bringen.
Q: Warum kam die Frau des Deutschen auf den Zehenspitzen ins Zimmer von Zeit zu Zeit?
A: Um dem Deutschen sein Essen zu bringen.
A = 25 B = 1 C = 430 D = (5).



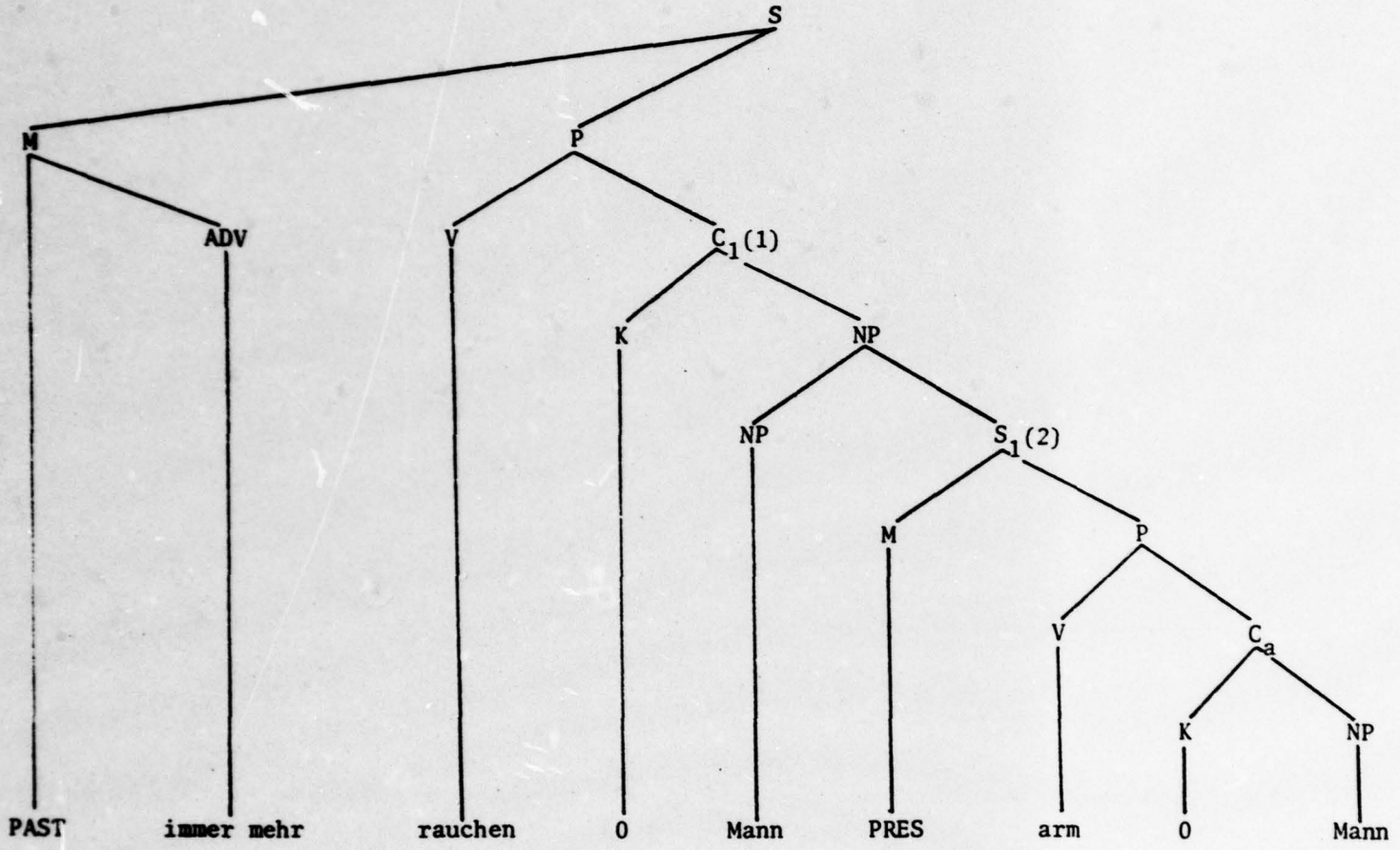
S: Am nächsten Abend kommt der Amerikaner mit einem vollen Notizbuch nach Hause.
Q: Womit kommt der Amerikaner am nächsten Abend nach Hause?
A: Mit einem vollen Notizbuch.
A = 27 B = 1 C = 331 D = (4).



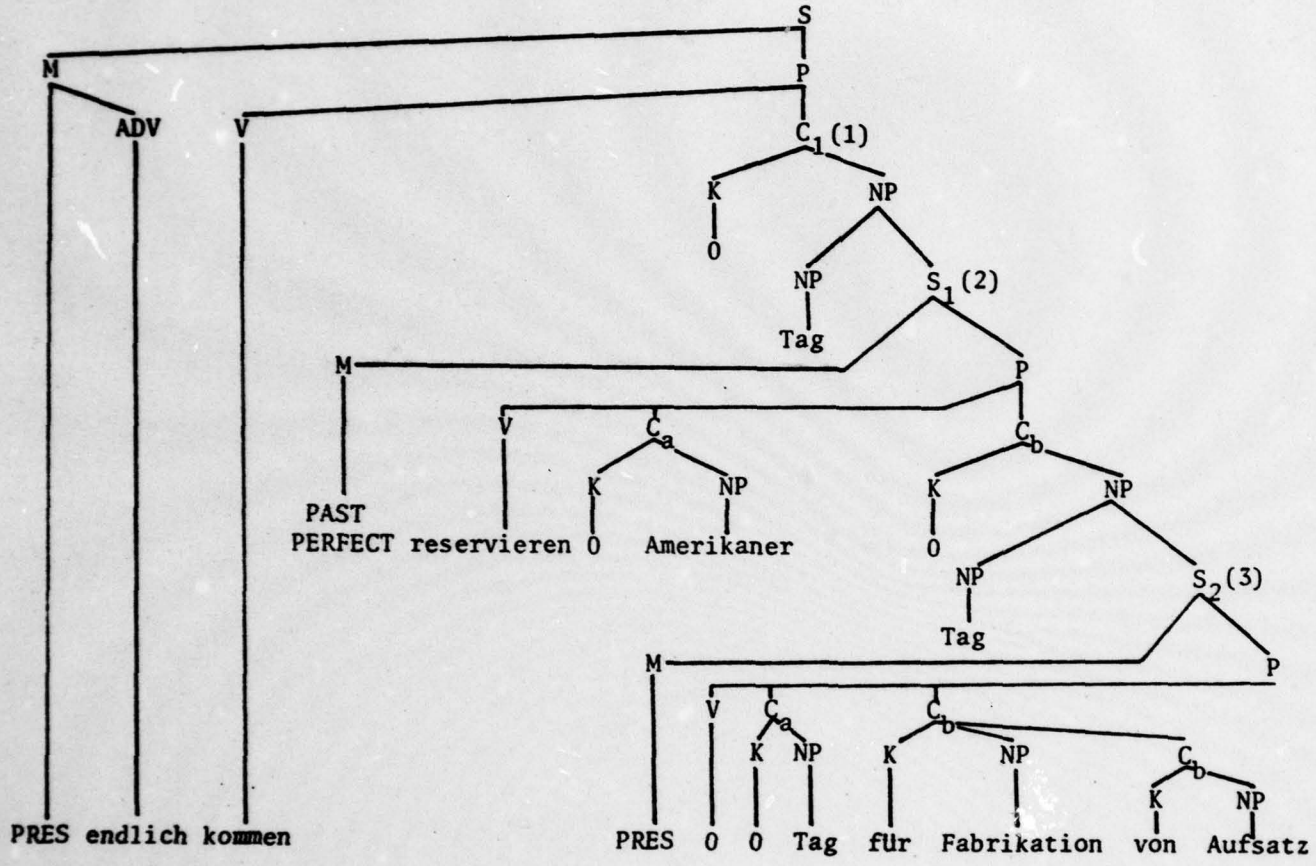
S: In England schrieb der Engländer ohne grosse Probleme und in klaren kurzen Worten alles, was der Engländer in der Heimat der Kamele gesehen hatte.
Q: Was alles schrieb der Engländer in England ohne grosse Probleme und in klaren kurzen Worten?
A: Alles, was er in der Heimat der Kamele gesehen hatte.
A = 30 B = 1 C = 145 D = (5).



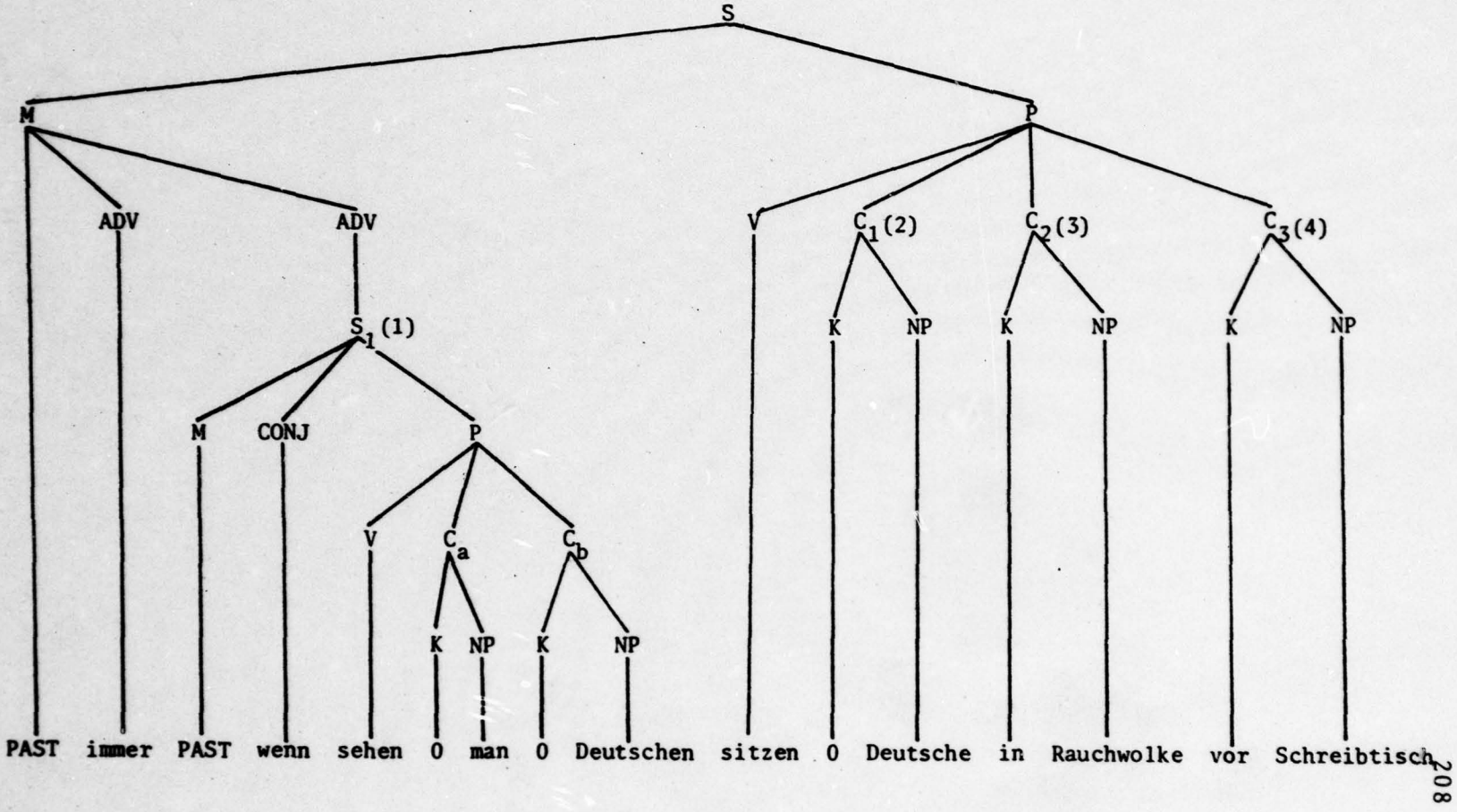
S: Der arme Mann <u>rauchte</u> immer mehr.
Q: Wer rauchte immer mehr?
A: Der arme Mann.
A = 31 B = 1 C = 452 D = (1).



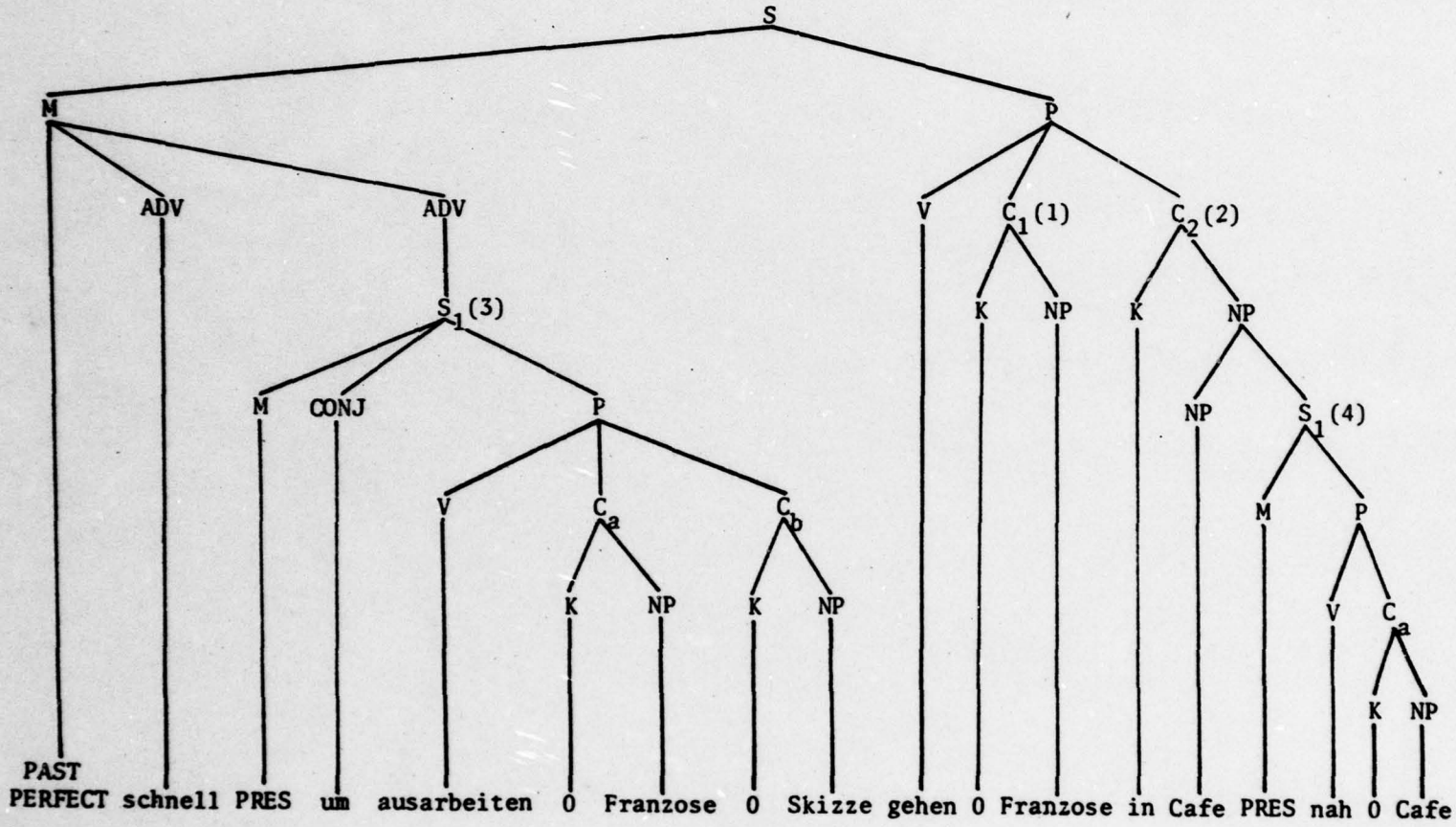
S: Endlich kommt der Tag, den der Amerikaner für die Fabrikation des Aufsatzes reserviert hatte.
Q: Was für ein Tag kommt endlich?
A: Der Tag, den der Amerikaner für die Fabrikation des Aufsatzes reserviert hatte.
A = 32 B = 1 C = 261 D = (2).



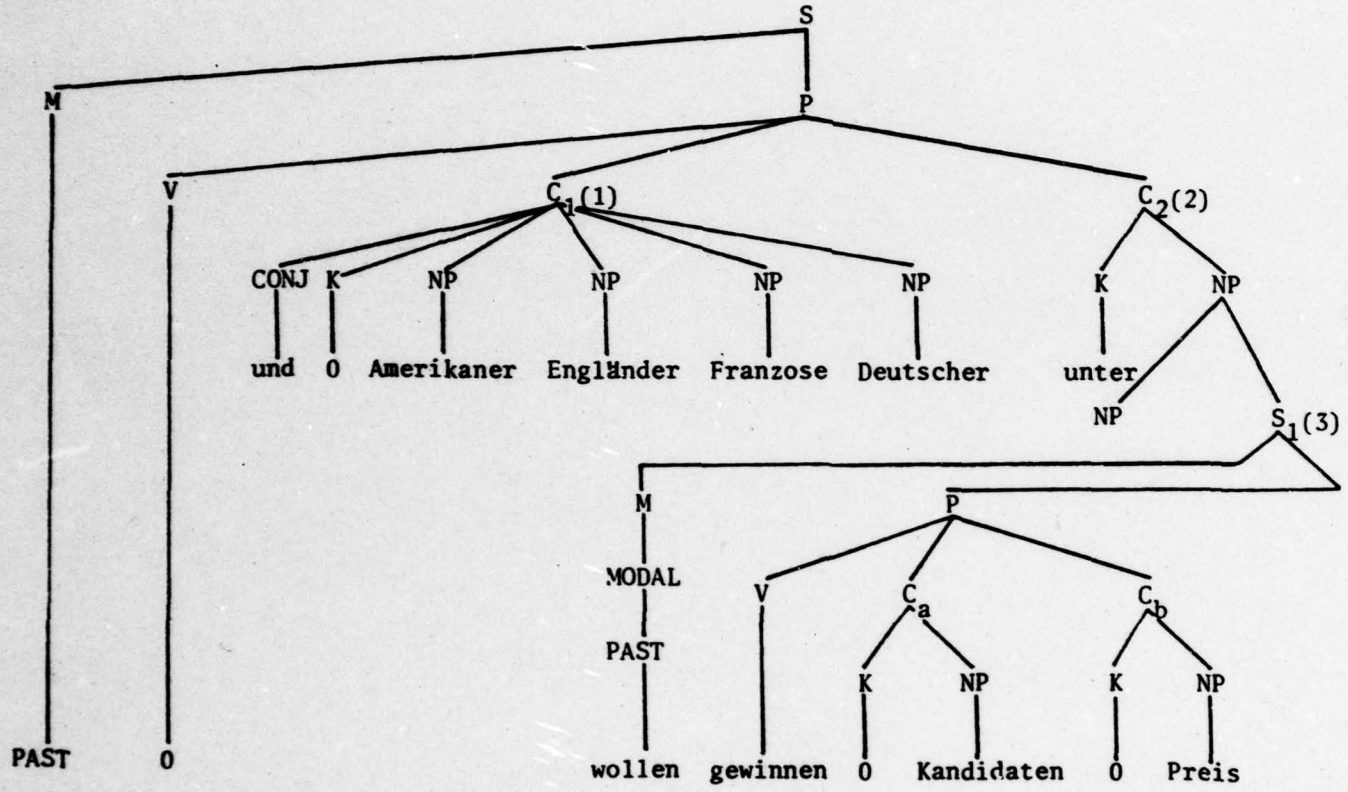
S: Wenn man den Deutschen sah, sass der Deutsche immer in einer Rauchwolke vor seinem Schreibtisch.
Q: Wovor sass der Deutsche immer in einer Rauchwolke, wenn man ihn sah?
A: Vor seinem Schreibtisch.
A = 33 B = 1 C = 473 D = (4).



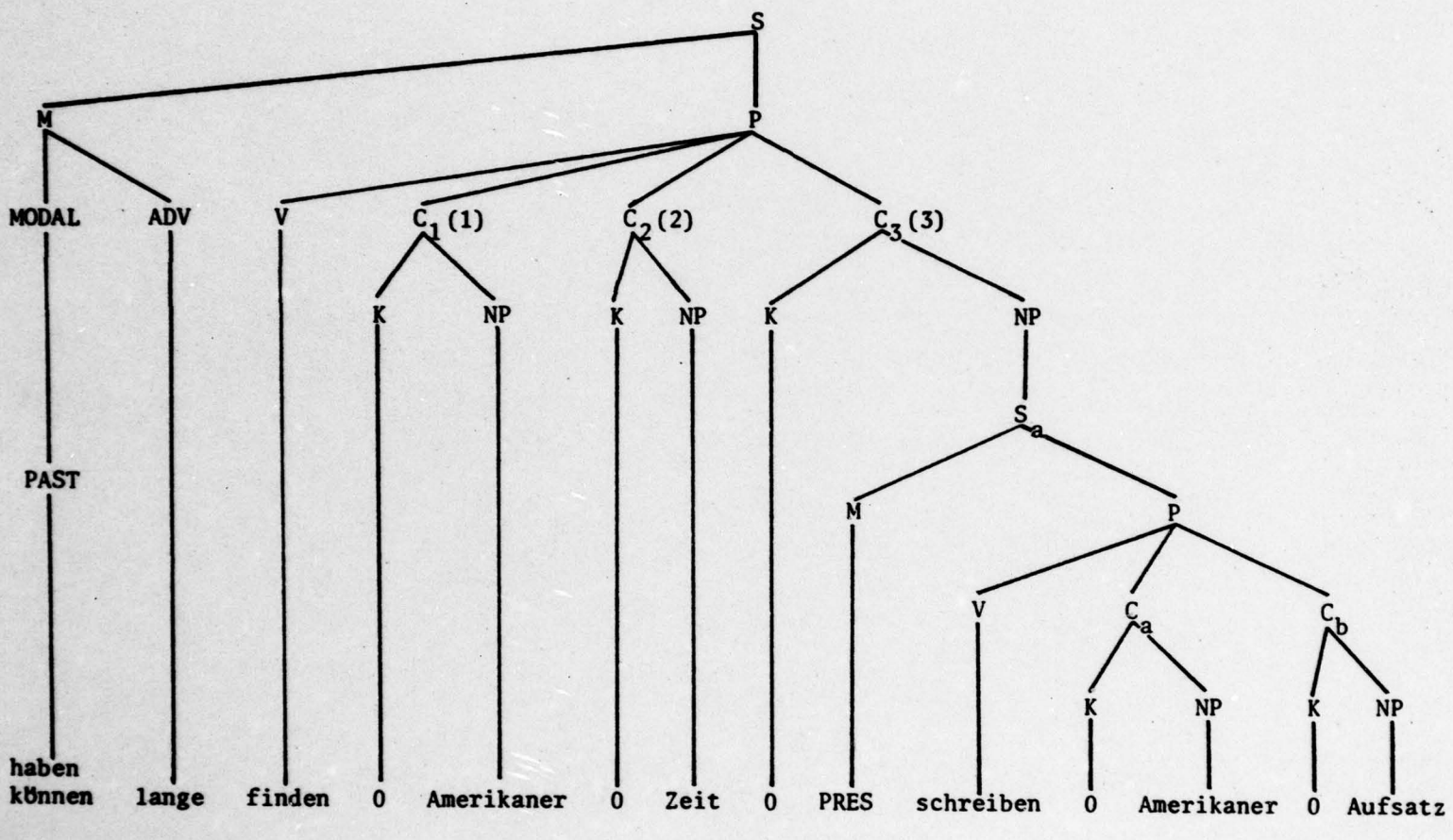
S: Schnell war der Franzose in ein nahes Cafe gegangen, um die Skizze auszuarbeiten.
Q: In was für ein Cafe war der Franzose schnell gegangen, um die Skizze auszuarbeiten?
A: In ein nahes Cafe.
A = 35 B = 1 C = 202 D = (4).



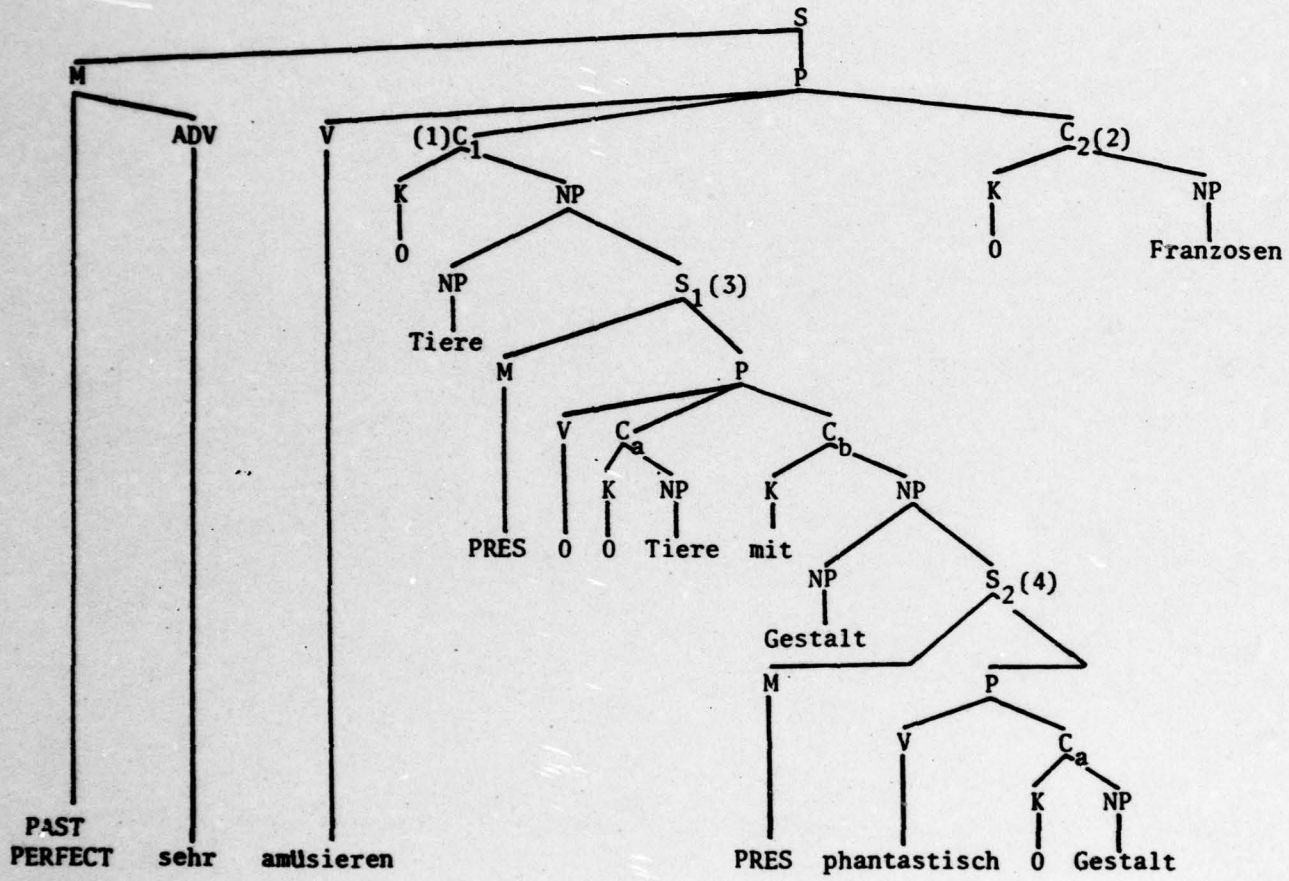
S: *Unter den Kandidaten, die den Preis gewinnen wollten, waren ein* Amerikaner, ein Engländer, ein Franzose und ein Deutscher.
Q: Unter was für Kandidaten waren ein Amerikaner, ein Engländer, ein Franzose und ein Deutscher?
A: Unter Kandidaten, die den Preis gewinnen wollten.
A = 39 B = 1 C = 27 D = (3).



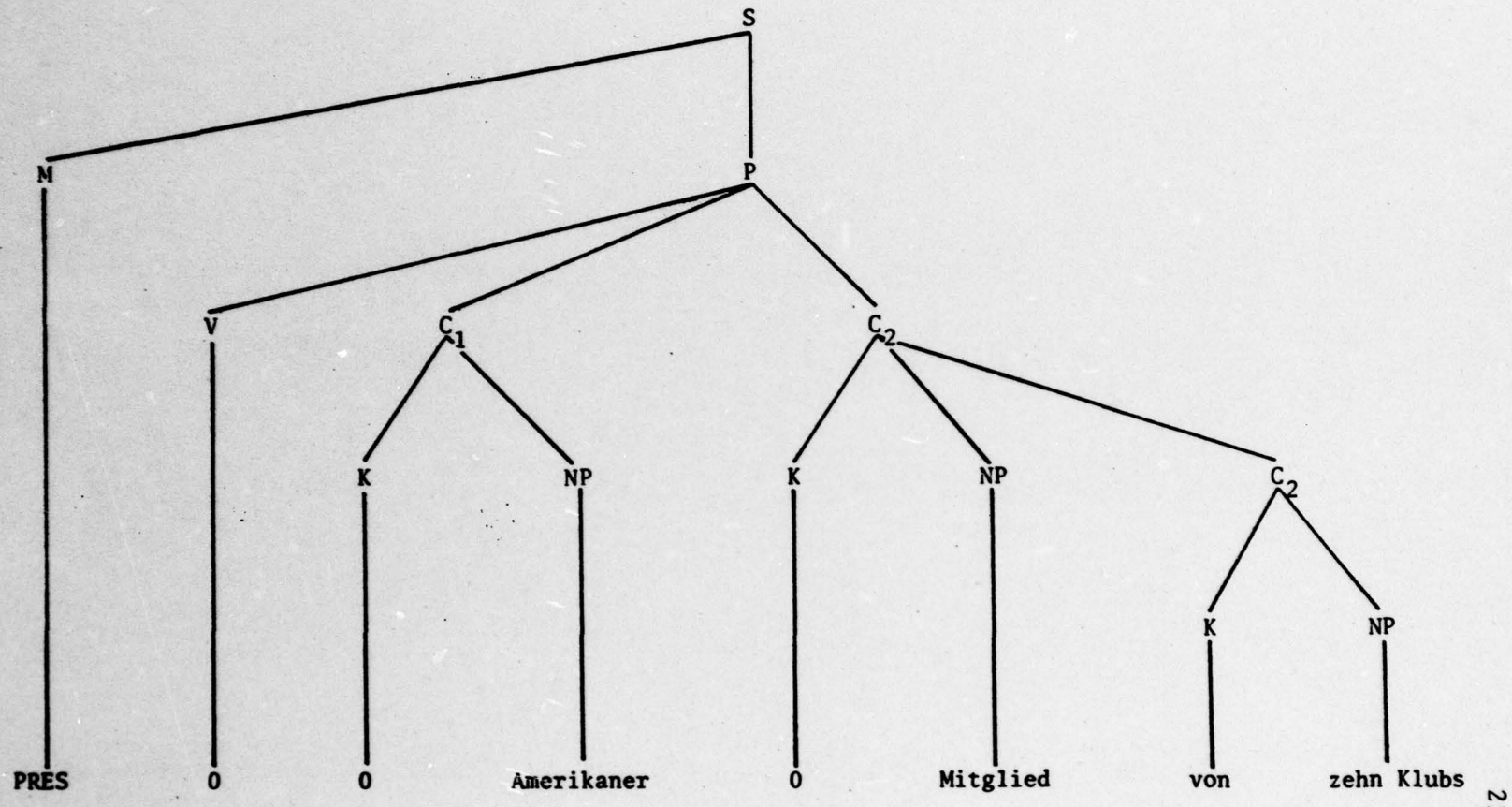
S: Der Amerikaner hatte lange keine Zeit finden können, den Aufsatz zu schreiben.
Q: Was zu tun, hatte der Amerikaner lange keine Zeit finden können?
A: Den Aufsatz zu schreiben.
A = 39 B = 2 C = 222 D = (3).



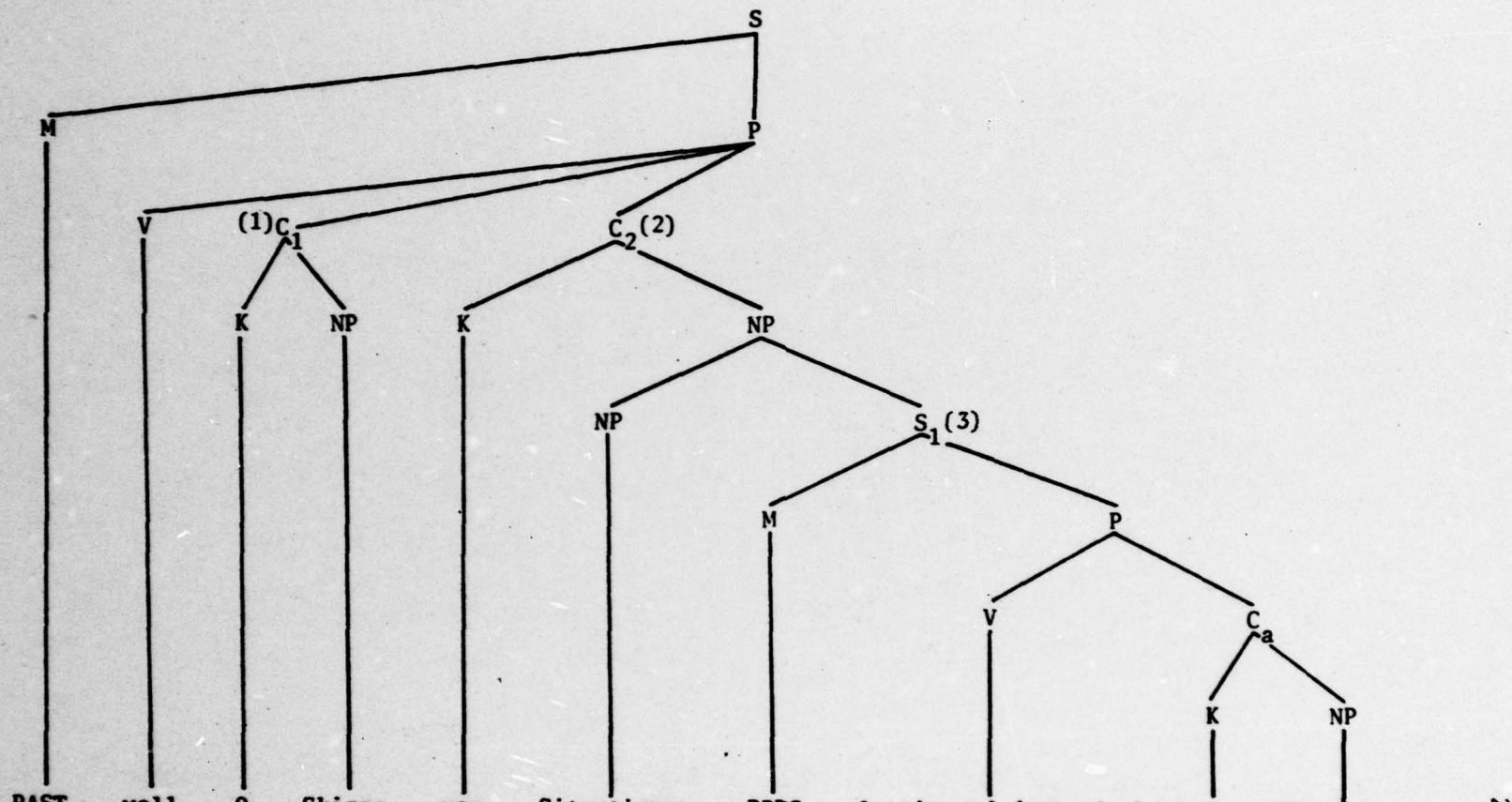
S: Die Tiere mit ihrer phantastischen Gestalt hatten den Franzosen sehr amüsiert.
Q: Wen hatten die Tiere mit ihrer phantastischen Gestalt sehr amüsiert?
A: Den Franzosen.
A = 44 B = 1 C = 183 D = (2).



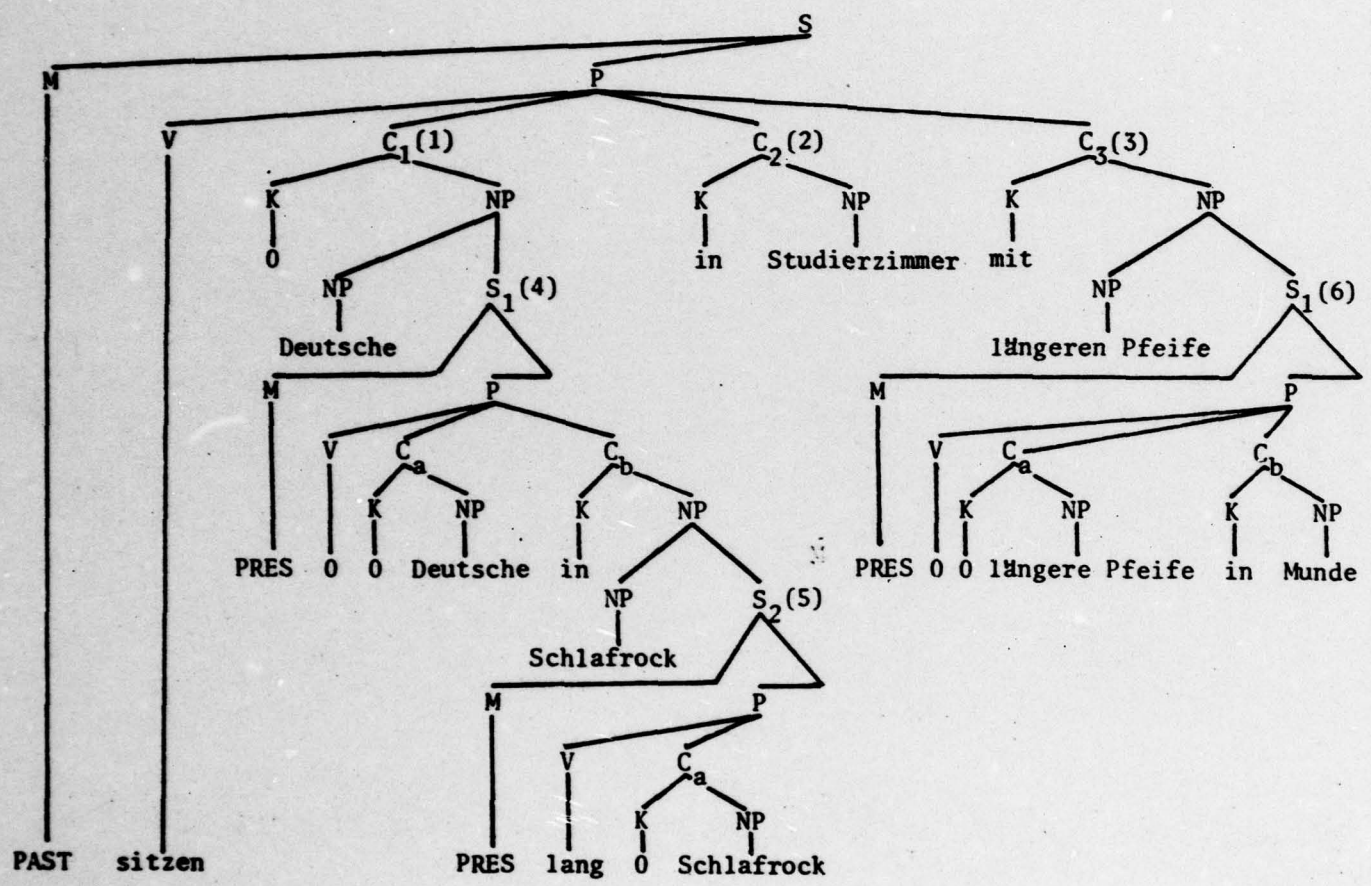
S: Der Amerikaner ist Mitglied von zehn Klubs.
Q: Wer ist Mitglied von zehn Klubs?
A: Der Amerikaner.
A = 45 B = 1 C = 244 D = (1).



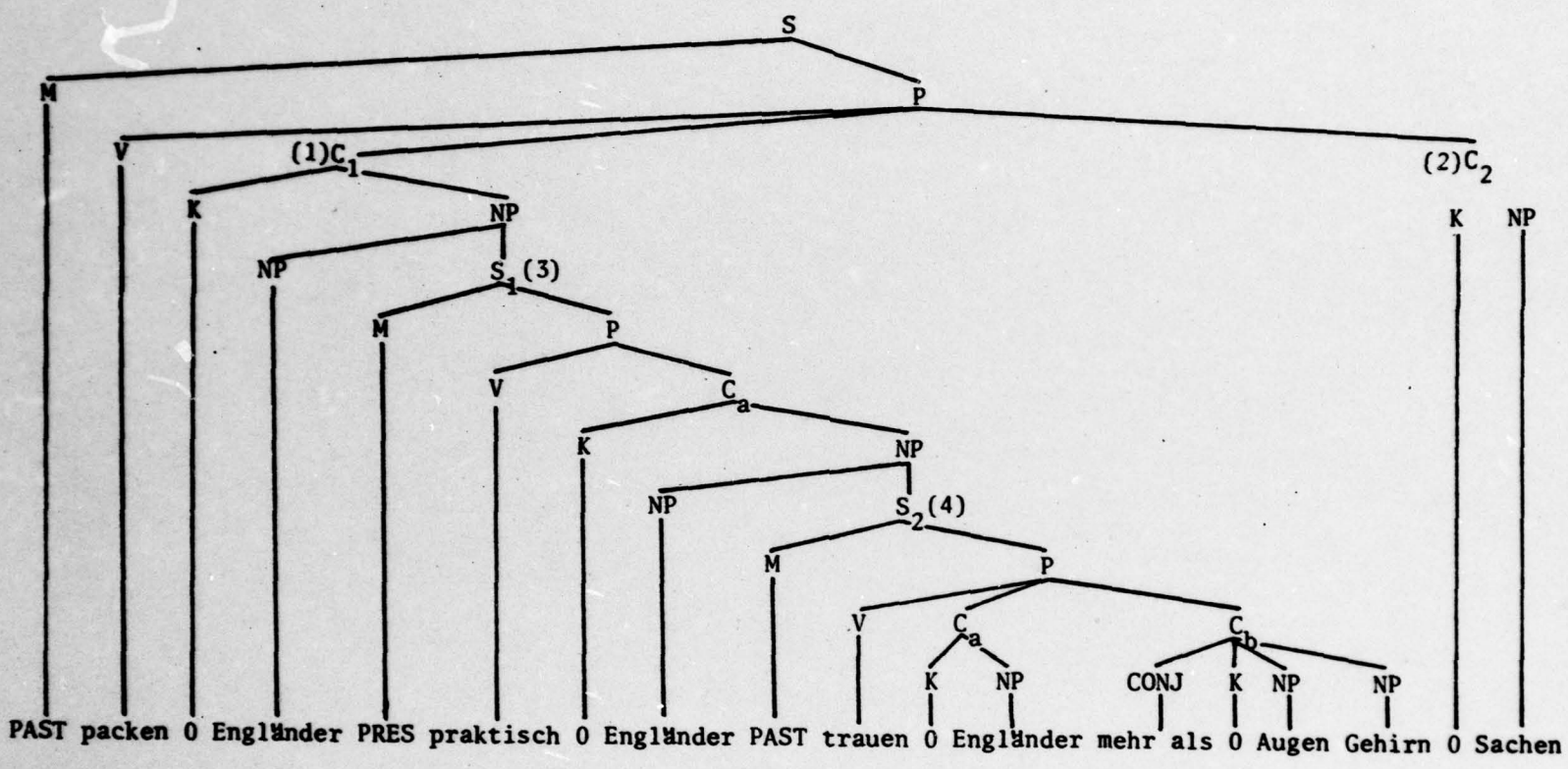
S: Die Skizze war voll lustiger und dramatischer Situationen.
Q: Was war voll lustiger und dramatischer Situationen?
A: Die Skizze.
A = 46 B = 1 C = 214 D = (1).



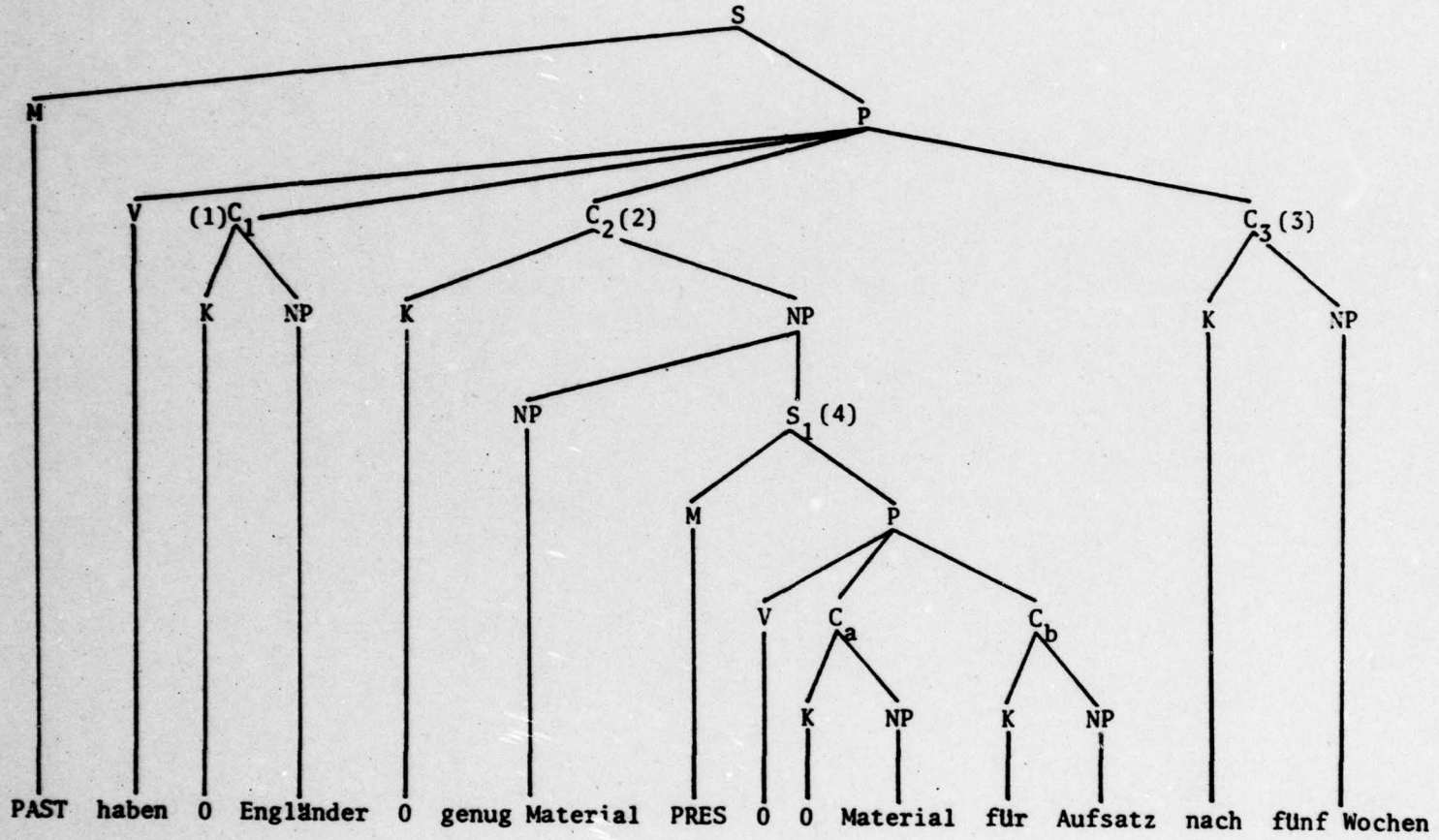
S: In seinem Studierzimmer sass der Deutsche im langen Schlafrock, mit einer noch längeren Pfeife im Munde.
Q: Wer sass in seinem Studierzimmer mit einer noch längeren Pfeife im Munde?
A: Der Deutsche.
A = 47 B = 1 C = 405 D = (1).



S: Der praktische Engländer, der seinen Augen mehr traute als seinem Gehirn, packte seine Sachen.
Q: Was für ein Engländer packte seine Sachen?
A: Der praktische Engländer.
A = 48 B = 1 C = 58 D = (3).



S: Nach fünf Wochen hatte der Engländer genug Material für den Aufsatz.
Q: Wer hatte nach fünf Wochen genug Material für den Aufsatz?
A: Der Engländer.
A = 49 B = 2 C = 119 D = (1).



BIBLIOGRAPHY

Alexander, H. W. "An Investigation of the Cloze Procedure as a Measuring Device Designed to Identify the Independent, Instruction, and Frustration Reading Levels of Pupils in the Intermediate Grades." Unpublished Ed.D. dissertation, University of Illinois, 1968.

Allen, E. D., and Valette, R. M. Modern Language Classroom Techniques. New York: Harcourt Brace Jovanovich, Inc., 1972.

Anderson, C. C. "A Factorial Analysis of Reading." British Journal of Educational Psychology, XIX (1949), 220-221.

Anderson, J. "An Investigation to Construct a Reference Scale by Means of Which the Difficulty Level of Primary Reading Books May Be Determined." Unpublished Master's Thesis, University of Queensland, 1965.

_____. Psycholinguistic Experiments in Foreign Language Testing. St. Lucia, Queensland, Australia: University of Queensland Press, 1976.

Barr, A. J., Goodnight, J. H., Sall, J. P., and Helwig, J. T. A User's Guide to SAS. Raleigh, North Carolina: SAS Institute INC., 1976.

Betts, E. A. Foundations of Reading Instruction. New York: American Book, 1946.

_____. "Readability: Its Application to the Elementary School." Journal of Educational Research, XLII (1949), 438-459.

Bickley, A. C., Ellington, B. J., and Bickley, R. T. "The Cloze Procedure." Journal of Reading Behavior, II No. 3 (1970), 232-249.

Bond, G. L., and Tinker, M. A. Reading Difficulties: Their Diagnosis and Correction. New York: Appleton-Century-Crofts, 1967.

Bormuth, J. R. "Cloze Tests as a Measure of Comprehension Ability and Readability." Unpublished Ph.D. dissertation, Indiana University, 1962.

_____. "Experimental Applications of Cloze Test." Proceedings of the International Reading Association, IX (1964), 303-306.

_____. "Validities of Grammatical and Semantic Classifications of Cloze Scores." Proceedings of the International Reading Association, X (1965), 283-286.

_____. "Cloze Readability Procedure." (EDRS: ED 010 983) 1967a.

_____. "The Implications and Use of Cloze Procedure in the Evaluation of Instructional Programs." (EDRS: ED 012 674) 1967b.

_____. "Comparable Cloze and Multiple Choice Comprehension Test Scores." Journal of Reading, X (1967c), 291-299.

_____. "Cloze Test Readability: Criterion Reference Scores." Journal of Educational Measurement, V (1968), 189-196.

_____. "Development of Readability Analysis." (EDRS: ED 029 166) 1969.

_____. On the Theory of Achievement Test Items. Chicago: The University of Chicago Press, 1970.

_____. "Development of Standards of Readability: Toward a Rational Criterion of Passage Performance. Final Report." (EDRS: ED 054 233) 1971.

_____. "Reading Literacy: Its Definition and Assessment." Reading Research Quarterly, IX No. 1 (1973-1974), 7-66.

_____. "The Cloze Procedure. Literacy in the Classroom." Help for the Reading Teacher. New Directions in Research. Edited by William D. Page. ERIC Clearinghouse on Reading and Communication Skills, National Institute of Education, March 1975.

_____, Manning, J., Carr, J., and Pearson, D. "Children's Comprehension of Between- and Within-Sentence Syntactic Structures." Journal of Educational Psychology, LXI (1970), 349-357.

Bowen, J. D. "A Tentative Measure of the Relative Control of English and Amharic by Eleventh Grade Ethiopian Students." Working Papers in TESL: UCLA, II (1969), 69-89.

Boyce, M. W. "A Comprehensive Bibliography of the Cloze Procedure." (EDRS: ED 099 830) 1974.

Briere, E. J., Clausing, G., Senko, D., and Purcell, E. "A Look at Cloze Testing Across Languages and Levels." Modern Language Journal, LXII (1978), 23-26.

Burkhart, K. H. "An Analysis of Reading Abilities." Journal of Educational Research, XXXVIII (1945), 430-439.

Carroll, J. B., Carton, A. S., and Wilds, C. P. "An Investigation of 'Cloze' Items in the Measurement of Achievement in Foreign Languages." (EDRS: ED 021 513) 1959.

Chall, J. S. Readability: An Appraisal of Research and Application. Columbus, Ohio: The Bureau of Educational Research, Ohio State University, 1958.

Chastain, K. The Development of Modern-Language Skills: Theory to Practice. Chicago: Rand McNally & Company, 1971.

Clausing, G., and Lange, D. L. "Cloze Testing and Language." (1977) Forthcoming in published form.

Coleman, E. B., and Miller, G. R. "A Measure of Information Gained During Prose Learning." Reading Research Quarterly, III (1968), 369-386.

Dale, E., and Chall, J. S. "A Formula for Predicting Readability." Educational Research Bulletin, XXVII (1948), 11-20.

Darnell, D. K. "The Development of an English Language Proficiency Test of Foreign Students Using a Clozentrophy Procedure: Final Report." (EDRS: ED 024 039) 1968.

Davis, F. B. "What Do Reading Tests Really Measure?" English Journal, XXXIII (1944), 100-107.

Draper, N. R., and Smith, H. Applied Regression Analysis. New York: John Wiley & Sons, Inc., 1966.

Dunn, O. J., and Clark, V. A. Applied Statistics: Analysis of Variance and Regression. New York: John Wiley & Sons, 1974.

Finn, P. J. "A Question Writing Algorithm for Written English." Unpublished Ph.D. dissertation, The University of Chicago, 1973.

Flesch, R. "Estimating the Comprehension Difficulty of Magazine Articles." Journal of General Psychology, XXVIII (1943), 63-80.

_____. "A New Readability Yardstick." Journal of Applied Psychology, XXXII (1948), 221-233.

Fletcher, J. E. "A Study of the Relationships between Ability to Use Context as an Aid in Reading and Other Verbal Abilities." Unpublished Ph.D. dissertation, University of Washington, 1955.

Fram, R. D. "A Review of the Literature Related to the Cloze Procedure." (EDRS: ED 075 785) 1972.

Friedman, M. M. "The Use of the Cloze Procedure for Improving the Reading Comprehension of Foreign Students at the University of Florida." Unpublished Ph.D. dissertation, University of Florida, 1964.

Gray, W. S. "Progress in the Study of Readability." Library Trends. Chicago: University of Chicago Press, (1937), 237-254.

Greene, F. P. "Modification of the Cloze Procedure and Changes in Reading Test Performances." Journal of Educational Measurement, II (1965), 213-217.

Hall, A. L. "A Readability Study in Intermediate College German." Unpublished Ed.D. dissertation, The University of Georgia, 1971.

Hanzeli, V. "The Effectiveness of Cloze Tests in Measuring French Students' Competence in an Academic Setting." (1977) Forthcoming in The French Review.

Harris, A. J. Effective Teaching of Reading. New York: David McKay, 1962.

Johnson, C. L. "Vocabulary Difficulty and Textbook Selection." Modern Language Journal, XI (1927), 290-297.

Jongsma, E. R. "The Cloze Procedure: A Survey of the Research." (EDRS: ED 050 893) 1971.

Jones, R. L. "Testing: A Vital Connection." The Language Connection: From the Classroom to the World. Edited by June K. Phillips. Skokie, Illinois: The National Textbook Company, 1977.

Kerlinger, F. N. Foundations of Behavioral Research. 2nd ed. New York: Holt, Rinehart and Winston, INC., 1973.

_____, and Pedhazur, E. J. Multiple Regression in Behavioral Research. New York: Holt, Rinehart and Winston, INC., 1973.

Kerr, A. H. "Cloze Procedure as a Measure of the Readability of French as a Second Language." Unpublished Bachelor's thesis, University of New England, 1968.

Klare, G. R. The Measurement of Readability. Ames, Iowa: Iowa State University Press, 1963.

_____, Sinaiko, H. W., and Stolurow, L. M. "The Cloze Procedure: A Convenient Readability Test for Training Materials and Translations." Paper P-660 (Arlington, Virginia: Institute for Defense Analyses Science and Technology Division) 1971.

Lorge, I. "Predicting Reading Difficulty of Selections for Children." Elementary English Review, VI (1939), 229-233.

_____. "Word Lists as Background for Communication." Teachers College Record, XLV (1944), 543-552.

_____. "Readability Formulae--An Evaluation." Elementary English, XXVI (1949), 86-95.

MacGinitie, W. H. "Contextual Constraint in English Prose Paragraphs." Journal of Psychology, LI (1961), 121-130.

Meier, H. Deutsche Sprachstatistik. Vol. II. 2nd ed. Hildesheim, Germany: Georg Olms Verlagsbuchhandlung, 1967.

Müller, M., and Wenckebach, C. Glück Auf. Boston: Ginn & Company, 1901.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., and Bent, D. H. Statistical Package for the Social Sciences. 2nd ed. New York: McGraw-Hill Book Company, 1975.

Oller, J. W. Jr. "Assessing Competence in ESL: Reading." TESOL Quarterly, VI (1972a), 313-323.

_____. "Scoring Methods and Difficulty Levels for Cloze Tests of Proficiency in English as a Second Language." Modern Language Journal, LVI No. 3 (1972b), 151-158.

_____. "Cloze Tests of Second Language Proficiency and What They Measure." Language Learning, XXIII No. 1 (1973), 105-118.

_____. "Research with Cloze Procedure in Measuring the Proficiency of Non-Native Speakers of English: An Annotated Bibliography." (EDRS: ED 104 154) 1975.

_____. "Evidence for a General Language Proficiency Factor: An Expectancy Grammar." Die Neueren Sprachen, XXV (1976a), 165-174.

_____. "Language Testing." A Survey of Applied Linguistics. Edited by Ronald Wardhaugh and H. Douglas Brown. Ann Arbor, Michigan: The University of Michigan Press, 1976b.

_____, and Inal, N. A Cloze Test of English Prepositions." TESOL Quarterly, V (1971), 315-326.

Phillips, J. K. "A Study of the Applicability of Task Analysis Methodology and Learning Hierarchies to Second Language Reading." Unpublished Ph.D. dissertation, The Ohio State University, 1974.

Potter, T. C. "A Taxonomy of Cloze Research, Part I: Readability and Reading Comprehension." (EDRS: ED 035 514) 1968.

Pressey, S. L., and Lively, B. "A Method for Measuring the Vocabulary of Textbooks." Educational Administration and Supervision, IX (October, 1923), 389-398.

Rankin, E. F. "An Evaluation of the Cloze Procedure as a Technique for Measuring Reading Comprehension." Unpublished Ph.D. dissertation, University of Michigan, 1957.

_____. "The Cloze Procedure--its Validity and Utility." Starting and Improving College Reading Programs. The Eighth Yearbook of the National Reading Conference. Edited by Oscar S. Causey and William Eller. Fort Worth, Texas: The Texas Christian University Press, 1959.

_____. "The Cloze Procedure--A Survey of Research." The Philosophical and Sociological Base of Reading. The Fourteenth Yearbook of the National Reading Conference. Edited by Eric L. Thurston and Lawrence E. Hafner. Milwaukee, Wisconsin: The National Reading Conference, Inc., 1965.

_____, and Culhane, J. W. "Comparable Cloze and Multiple-Choice Comprehension Test Scores." Journal of Reading, XIII (1969), 193-198.

Ransom, P. E. "Determining Reading Levels of Elementary School Children by Cloze Testing." Proceedings of the International Reading Association, XII Part 1 (1968), 477-482.

Ruddell, R. B. "An Investigation of the Effects of the Similarity of Oral and Written Patterns of Language Structure on Reading Comprehension." Unpublished Ph.D. dissertation, University of Indiana, 1963.

Scherer, G. "Reading for Meaning." Northeast Conference on the Teaching of Foreign Languages, (1963), 22-6).

Shiba, S. "A Study of Reliability Measurement--Application of Cloze Procedure to the Japanese Language." Japanese Journal of Psychology, XXVIII (1957), 67-69.

Spaulding, S. J. "Reading Difficulty of Materials Used in Spanish Ph.D. Language Examinations." Unpublished Master's thesis, The Ohio State University, 1950.

_____. "A Spanish Readability Formula." Modern Language Journal, XL (1956), 433-441.

Stockwell, R. P., Schachter, P., and Partee, B. H. Integration of Transformational Theories on English Syntax. Contract No. AF 19(628)-6007 by the University of California, Los Angeleles: UCLA, 1968.

Taylor, W. L. "Cloze Procedure: A New Tool for Measuring Readability." Journalism Quarterly, XXX (1953), 415-433.

_____. "K. M. Readers Lend Hand to Science: 'Cloze' Method Works in Written Korean and May Serve As a Tool for Korean Language Reform." Korean Messanger, III (1954), 4-5.

_____. "Recent Developments in the Use of 'Cloze Procedure.'" Journalism Quarterly, XXXIII (1956), 42-48, 99.

Teitelbaum, H., Edwards, A., and Hudson, A. "Ethnic Attitudes and the Acquisition of Spanish as a Second Language." Language Learning, XXV (1975), 255-266.

Tharp, J. "The Measurement of Vocabulary Difficulty." Modern Language Journal, XXIV (1939), 169-178.

Thorndike, E. L. "Reading and Reasoning: A Study of Mistakes in Paragraph Reading." Journal of Educational Psychology, VIII (1917), 323-332.

Walter, R. B. "Historical Overview of the Cloze Procedure." (EDRS: ED 094 337) 1974.

Weaver, W. W. "An Examination of Some Differences in Oral and Written Language Using the Cloze Procedure." Unpublished Ph.D. dissertation, University of Georgia, 1961.